PROBLEMS IN THE OPTIMAL CONTROL OF FINITE AND INFINITE DIMENSIONAL LINEAR SYSTEMS

K.T.PARKER

A thesis presented for the
degree of Doctor of Philosophy,
The University of Warwick,
November 1975.

PREFACE

I should like to thank my family and friends for all the help and encouragement given to me during the time I have been working on this thesis, especially my wife, Anne, who has had to put up with it taking over our home and to Miss Linda China for her invaluable help with the typing. Above all, though, I must say how grateful I am to my supervisor, Dr. Tony Pritchard, not only for his help while I was at Warwick but also for the patience and understanding he has shown since. Also I must express my thanks to my examiners for the first submission of this thesis, Professors Parks and Collins, for their helpful suggestions and finally to the Science Research Council who provided me with a postgraduate grant. I feel I must also say that my present employers, the Department of the Environment, have been very generous in giving me special leave in order to finish the thesis.

I worked in close collaboration with my supervisor, Dr. Pritchard, so it is not easy to decide to whom each piece of work should be ascribed. Chapter 1 is my account of his work on optimal control theory, though Section 3 was the result of much mutual discussion. The method for finding bounds on the cost, Chapter 2, Section 4, is basically Dr. Pritchard's work with me carrying out some of the extra details. All the computing here, as in all the thesis, is completely my work. Chapter 3, Section 3, is also collaborative and has been published previously, Pritchard and Parker(1971), as has the work on the bounds for the cost, Pritchard and Parker(1974), and the design of PID controllers for finite dimensional systems, Parker(1972).

Kim Parker, the Department of Engineering, the University of Warwick.
November 1975.

ABSTRACT

A review of optimal control theory for linear systems with quadratic cost functions is presented. Some of the theoretical and practical limitations are discussed with special reference to distributed parameter systems. First a procedure is described for finding the optimal control by constructing a sequence of controllers that converges to the optimal; this method is valid for systems of infinite dimension provided that the operators in the state differential equation satisfy certain conditions. The proof is carried out both for the finite and infinite time interval and the connection is shown with the Riccati equation. The main problem in implementation is that one needs complete knowledge of the state at all times in order to build the optimal controller, this is almost certainly impossible for distributed parameter systems. When the state cannot be measured completely it is proved that an optimal control is realisable for time invariant finite dimensional systems. The problems of finding this control are then investigated and computational methods discussed. If the optimal control with complete knowledge of the state cannot be implemented, a method is presented whereby one can find bounds on the possible increase in the value of the cost function arising from the use of some sub-optimal control; several examples are considered. The constrained optimal control depends on the initial state and new optimisation criteria must be put forward to deal with the case in which the initial state is

unknown; the most common consist of minimising the cost that can result from the worst initial state. It is then shown how the controllers designed according to these criteria may be improved by using one's limited observation at time zero to place some constraints on the initial state. The Liapunov matrix equation plays an important part in calculating the cost of any control so reducing the computational effort in its solution is useful. It is shown how this can be done and it is of special relevance for distributed parameter systems with their states expressed as an infinite series of eigenfunctions; the results are applied to a diffusion equation example. Finally, it is shown how optimal control theory may be applied to the design of proportional-integral-derivative controllers. This is done from two standpoints and the resulting controllers are shown to be identical, though the second method of proof is valid for infinite dimensional systems. The results are then applied to a simple example and to a distributed population dynamics system. The practicality of the methods of the thesis are applied to a system with realistic parameters; recommendations are made as to the best approaches.

--- oOo ---

CONTENTS

INTRODUCTION

This thesis is concerned with the optimal control of linear systems with quadratic cost functions and some of the problems that can arise in its implementation. We shall consider especially the aspects related to infinite dimensional systems, both the optimal control theory itself and the special difficulties involved in its practical application.

The work is restricted to the so called linear quadratic problem as this represents a practicable approach to the control of many physical systems. Obviously any real system, if analysed in sufficient detail, will be non linear but there is a trade off in any mathematical model between accuracy of description and feasibility of obtaining useful results without undue difficulty. A well thought out linear analysis can often show important characteristics of a system and for these reasons a linear model is often a good starting point for working out the response to different inputs and controls. The object of controlling a system is to drive it into some desired state and, if one is only considering perturbations about that state, a linearised model will give an adequate description provided the perturbations are not too large. The reasons for associating a quadratic cost function with the problem are not so clear cut. The cost function can be considered as a penalty, arising from any control, that must be minimised, so including a quadratic term in the deviation of the state from its desired value

is intuitively reasonable, there should be a positive penalty for both over and undershooting. Similarly there ought to be a term containing the actual control action and this too should be positive irrespective of the sign of the control variables. However, there is no absolute reason for choosing quadratic, rather than say quartic or modulus, functions; it is more a matter of expediency. The great advantage of using a quadratic cost function with linear systems is that it gives rise to a linear feedback controller, something that is very desirable. It would not be particularly useful to set up a linear model only to find that the optimal control thus derived was non-linear, so nullifying the work that went into finding a good linearisation. Also the classical methods of designing control systems are all based on the idea of linear negative feedback, therefore the "modern" control theory is not at odds with the "classical". So, all in all, the linear quadratic approach is a reasonable one to take for many practical systems, though obviously one must exercise caution and not "bend" the physics and engineering of the system too much in order to fit in with elegant mathematical theory for its own sake.

The main starting point for this work was a paper by Athans (1970) who pointed out the impossibility of building an optimal controller for linear distributed parameter systems, that is those described by partial differential equations. He suggested that since a finite number of spatially discrete sensors would have to be used for

measurement some account of the complication of the sensing apparatus should be taken in the cost function. This led to the analysis of some particular examples, for instance that in chapter 3, section 3, Pritchard & Mayhew (1971) and Parker (1970). However, a more general approach had to be made and theory developed that would be appropriate to this problem.

The characteristic property of distributed parameter systems is that they are of infinite dimension. This means that if the concept of a state space, Ogata (1967), is applied to the system, one needs an infinite number of state variables to describe the system completely. To clarify this idea consider the flow of heat along a thin homogenous bar. It is well known, Brogan (1968), that the equation governing this system is the parabolic partial differential diffusion equation $\frac{\partial z}{\partial t} = \frac{\partial^2 z}{\partial x^2}$ where z represents the temperature at time t at a point x from the end of the bar. In this case the temperature distribution along the bar at any instant is the state of the system and in order to measure it completely the temperature has to be observed at every point on the bar. Hence there are an infinite number of components to the state. As in systems of finite dimensions, Ogata (1967), there is not one unique set of state variables; it is possible to carry out a linear transformation so that the state differential equations are in a more convenient form. One way of doing this in the example mentioned is to express the temperature distribution as a suitable Fourier series, then the coefficients of this

series, which are infinite in number, can be used as new state variables; this is done in some of the specific examples considered in the later chapters. However the state is expressed it is obvious that there will be great difficulty in measuring it exactly.

Optimal control theory has been very well developed for finite dimensional systems, Bellman (1967), Kalman (1960), Pontryagin et al (1962), Athans & Falb (1966) and some of this has been carried over formally to distributed parameter systems, Wang (1964), Kim & Erzberger (1967), Brogan (1968). Unfortunately these methods are not rigorous as they make assumptions about the nature of the solutions to the partial differential equations that are not justified; for example, the state may not be continuously differentiable with respect to time. Nevertheless, there are many situations in which they give useful results. Lions (1971) and Pritchard (1972) by framing the problem in the appropriate spaces and imposing certain conditions on the operators in the equations have developed optimal control theory that is valid for a wide range of infinite dimensional systems. Moreover, if further restrictions are made it can be shown that their methods give the same results as those derived in a purely formal manner.

The optimal control theory for the linear quadratic problem yields a linear feedback controller that results in the cost function having a global minimum; another advantage is that the optimal control turns out to be independent of the initial state of the system. However,

next section demonstrates the connection between this optimal control and the standard Riccati equation. Finally the results for the optimal control are extended to cover the case of the infinite interval. This can only be done if it is possible to construct controls that yield a finite value for the cost.

Chapter 2 deals with the limitations of the optimal control and the problems of observing the complete state are considered in more detail. The existence of an optimal control when one has incomplete information about the state is discussed and it is proved that the finite dimensional time invariant system does have a realisable optimal control provided that it is possible to find some controller that gives a finite cost. The next step is to consider ways of actually calculating the optimal feedback when the state cannot be measured completely, the so called constrained optimal control problem. The method of Jameson (1970) is presented and the difficulties arising discussed. He derives the conditions for the cost to have a stationary value, but obviously this is only a necessary and not a sufficient condition for the cost to be a minimum. An iterative method, the fractional step algorithm, is derived from these equations and it is shown that for small steps there is a guaranteed reduction in the value of the cost function. The final section of this chapter presents a method for finding bounds on the cost function which are useful as a relatively easy way of comparing two controls. The proof is valid for all systems that satisfy the

conditions laid down in chapter 1 and so a wide variety of infinite dimensional systems can be considered as well. Finally the method is applied to some specific examples to find the bounds on their cost functions.

Chapter 3 is primarily concerned with the fact that when the observation of the state is restricted the control that minimises the cost function depends on the initial state. It is shown how, if the probability density function of the initial state is known, it is possible to find a simple expression for the expected value of the cost and in some cases the optimal feedback can be found by modifying Jameson's equations. The "worst case" criteria are discussed wherein the control that minimises the greatest cost that can occur for some unknown initial state is found. A method of improving this method is presented in which the fact that one has limited knowledge of the state at time zero is used, this is then applied to an example. The Liapunov matrix equation which has to be solved in order to calculate the cost is very important in all the analysis considered in this thesis; any method of simplifying its solution is therefore of great use. A way of reducing the computational effort that has special relevance for distributed parameter systems is explained and applied to a diffusion equation example.

In chapter 4 we lock at some of the differences and points of contact between modern and classical control theory. It is then shown how optimal control theory can be used to

design a proportional-integral-derivative type controller, something that is usually considered to be in the classical realm. The proof is carried out from two points of view, both leading to the same controller. The second of these is valid for infinite dimensional systems satisfying the conditions specified in chapter 1. The procedure is first applied to a simple second order example and then to a distributed parameter system. The latter consists of finding an optimal culling policy for red deer; a considerably simplified discretisation is used for the actual computation.

In the final chapter the methods developed in the thesis are applied to a third order water flow system in which the parameters have been made as realistic as possible. The choice of the cost function is considered in some detail and the optimal control derived. Various numerical procedures are tested for the constrained optimal control problem; none turns out to be completely satisfactory except the fractional step algorithm.

Some mention must be made here of the terminology and notation. Every effort has been made to use consistent conventions, but often different terms are used to define the same thing, this is done in order to reduce repetition of certain common phrases. For example, cost function is also referred to as performance index and the fact that the state cannot be measured completely is described at various times by the expressions, partially observed system,

constrained feedback and optimal control, incomplete information, incomplete knowledge, not measuring the state completely and so forth. Conversely their opposites are used when describing the optimal control when one has complete knowledge of the state. It is hoped that what is meant is clear in the context and the variation helps the flow of the argument.

The state of an infinite dimensional system is always referred to as $z(t)$ while the state vector of a system with finite dimension is designated by $x(t)$. A prime denotes the transpose of a vector or matrix and a star the adjoint of an operator. The latter, though, is sometimes used as an indicator for the optimal control, for example $x^*(t)$ is the optimal trajectory of the state vector. Finally, the notation $t \in [0,T]$ indicates that t lies on the interval between 0 and T and includes both end points; if an end point is to be excluded a round bracket is used, $t \in [0,T)$ implies that $0 \leq t < T$.

--- oOo ---

CHAPTER 1

THE THEORY OF OPTIMAL CONTROL

Section 1. The formulation of the problem.

The theory of optimal control has been very well developed for finite dimensional systems, Kalman(1960), Bellman(1957), Athans & Falb(1966), and many of the results can be formally carried over to systems of infinite dimension, Wang(1964), Kim & Erzberger(1967). However,this step involves making assumptions that cannot always be justified. One approach to this problem has been made by Lions(1971) who assumes that the differential operator, A, can be associated with a bilinear form; he then considers a decoupled set of equations formed by introducing the adjoint state. In this chapter we shall give an account of a method presented by Pritchard (1972) in which the conditions placed on A are less restrictive than those necessary in Lions' work.

The state of the system, z(t), and the control input, u(t), are taken to be elements of the Hilbert spaces H and U respectively. The inner products and norms in these spaces will be designated by $\langle\ ,\rangle_H$ $\langle\ ,\rangle$, $\|\ \|_H$ and $\|\ \|$. In many cases the system will be given as a partial differential equation with z and u defined on a spatial domain Ω with boundary $\partial\Omega$. In this case it will be assumed that a closed linear operator A, defined on a dense domain $D(A)\subseteq H$, can be defined from the formal differential operators by assuring that z is in the appropriate space. The system equations are then expressed as

$$\frac{dz(t)}{dt} = Az(t) + B\,u(t)$$
$$z(0) = z_0 \qquad (1.1.1)$$

where B is a linear bounded operator, $U \to H$, which is also derived from the formal differential operators of the origihal system.

However, if there is control action at the boundary the operator B will be unbounded. In this case the following analysis is not valid, though in many systems it is possible to find the optimal control by the formal methods mentioned before, see Pritchard(1972).

A control u(t) will be said to be admissible if it is measurable on $[0,T]$ in such a way that

$$\int_0^T \|u(t)\|^2 dt < \infty .$$

The first consideration must be to define exactly what is meant by a solution of (1.1.1). z(t) is said to be a strict solution if it is differentiable and satisfies (1.1.1) everywhere. The solution is weak if z(t) only satisfies

$$\int_0^T \left[\phi(t)\langle Az + Bu, y\rangle_H + \frac{d\phi}{dt}(t)\langle z, y\rangle_H\right] dt = 0 \qquad (1.1.2)$$

and $\langle z(0) - z_0, y\rangle_H = 0$, $y \in D(A)$, $\phi(t)$ sufficiently smooth with compact support in $[0,T]$

It will be assumed that A generates a strongly continuous semigroup T_t, $t>0$, of linear bounded operators satisfying

$$\frac{dT_t z}{dt} = AT_t z , \qquad T_{t+s} = T_t T_s \qquad (1.1.3)$$

$T_0 = I$, the identity operator,

where $z(t) \in D(A)$. The conditions for this to be so are given in Yosida(1965). The adjoint operator, A^*, will also generate a strongly continuous semigroup T_t^* and

$$\|T_t\|_H = \|T_t^*\|_H \leq Me^{\beta t}$$

for some $M>0$, $\beta<\infty$. The solution to (1.1.1) can now be written in terms of T_t thus

$$z(t) = T_t z_0 + \int_0^t T_{t-s} B\ u(s)ds \qquad (1.1.4)$$

If a strong solution is not known to exist then as long as it is possible to prove that A generates a strongly

continuous semigroup, (1.1.4) will be said to be the mild solution of (1.1.1).

It will also become necessary to consider the equation

$$\frac{dz(t)}{dt} = F(t)z(t) + B\,u(t) \tag{1.1.5}$$

where $F(t) = A + E(t)$. For the case under consideration we have that $E(t)$ is a bounded linear transformation for every t and is strongly continuous in t on every finite interval in $[0,\infty)$, and since $B\,u(t)$ is strongly continuously differentiable on $[0,T]$, Pritchard & Curtain (1975) have shown that there is a mild solution to (1.1.5) given by

$$z(t) = S(t,0)z_0 + \int_0^t S(t,\sigma)\,B\,u(\sigma)\,d\sigma \tag{1.1.6}$$

$z_0 \in H$.

Here $S(t,\sigma)$ is an evolution operator satisfying the following conditions

$$S(t,\sigma)x = T_{t-\sigma}x + \int_\sigma^t T_{t-\rho}E(\rho)S(\rho,\sigma)x\,d\rho$$

$x \in H$

$$S(t,\sigma)S(\sigma,\tau) = S(t,\tau) \quad t \geq \sigma \geq \tau\,, \tag{1.1.7}$$

$$S(t,t) = I\,, \quad \|S(t,\sigma)\|_H < Me^{\alpha(t-\sigma)},\ M>0\,,\ \alpha < \infty.$$

We now associate a cost functional with the problem that will be taken to be of the standard quadratic form

$$J(u) = \langle z(T), Gz(T)\rangle_H + \int_0^T \left[\langle z(\sigma),\ Q\ z(\sigma)\rangle_H + \langle u(\sigma),\ R\ u(\sigma)\rangle_U\right]d\sigma\,. \tag{1.1.8}$$

Q and G are bounded self-adjoint positive semidefinite operators on H, while R is a bounded self-adjoint positive definite operator on U.

The objective of the following work is to show that there exists $u^* \in U$ such that

$$\inf_{u\in U} J(u) = J(u^*)$$

and to find a way of determining u^*. This is done by finding a sequence of admissible controls

$$u_i(t) = K_i(t)z(t)$$

such that

$$J(u_{i+1}) \leqslant J(u_i)$$

for all i, and that

$$u_i \to u^* , \quad J(u_i) \to J(u^*) \qquad \text{as } i \to \infty .$$

Firstly control on the finite interval $[0,T]$ will be considered. Assume that we have some control, $u_i(t)$, and that it is perturbed by $\bar{u}(t)$ so that

$$u(t) = u_i(t) + \bar{u}(t) = K_i(t)z(t) + \bar{u}(t) .$$

Therefore the equation governing the system can be written symbolically

$$\frac{dz}{dt}(t) = A_i(t)z(t) + B\,\bar{u}(t) \qquad (1.1.9)$$

where

$$A_i(t) = A + B\,K_i(t) . \qquad (1.1.10)$$

Having shown which variables depend on time, for the sake of clarity, the (t) indicating dependence on time will be suppressed where it is possible to do so without causing ambiguity. The unique solution to (1.1.9) is given by equation (1.1.6), that is

$$z(t) = S_i(t,0)z_0 + \int_0^t S_i(t,\sigma)\,B\,u(\sigma)d\sigma \qquad (1.1.11)$$

where $S_i(t,\sigma)$ is the evolution operator derived from $A_i(t)$. In his paper (1972) Pritchard proves what will be called:

Theorem I

If

$$P_i(t)x = S_i^*(T,t)GS_i(T,t)x + \int_t^T S_i^*(\sigma,t)Q_i(\sigma)S_i(\sigma,t)\,d\sigma$$
$$x \in H \qquad (1.1.12)$$

then

$$\langle z(t),P_i(t)z(t)\rangle_H = \langle z(T),Gz(T)\rangle_H + \int_t^T \left[\langle z,Q_i z\rangle_H - 2\langle z,P_i B\bar{u}\rangle_H\right]d\sigma$$

where

$$Q_i = Q + K_i^* R K_i$$

and * denotes the adjoint operator. This is proved by substituting directly for z and P_i and then using the properties of $S_i(t,\sigma)$ together with some integral identities; the procedure is straightforward but involves a large amount of manipulation and the application of some well known theories in analysis. From theorem I we can show how it is possible to generate a sequence of controls of decreasing cost, this is done in

Theorem II.

Let

$$K_i = -R B^* P_{i+1} \quad , \quad K_0 = 0, \tag{1.1.13}$$

if

$$u_i = K_i z$$

then (a) $J(u_i) = \langle z_0, P_i(0) z_0 \rangle_H$

(b) $\langle z(t), P_{i+1}(t) z(t) \rangle_H \leq \langle z(t), P_i(t) z(t) \rangle_H$.

Proof: (a), in theorem I put $\bar{u} = 0$, then

$$\langle z(t) P_i(t) z(t) \rangle_H = \langle z(T), Gz(T) \rangle_H + \int_t^T [\langle z, Qz \rangle_H + \langle u_i, Ru_i \rangle_U] d\sigma .$$

If $t = 0$ the right hand side equals $J(u_i)$ by definition from (1.1.8), so

$$\langle z_0, P_i(0) z_0 \rangle_H = J(u_i) .$$

(b) set

$$u = K_i z + \bar{u}$$

then, using (a)

$$\langle z(t), P_i(t) z(t) \rangle_H - \langle z(t), P_{i+1}(t) z(t) \rangle_H = \langle z_0, P_i(0) z_0 \rangle_H$$
$$- \int_t^T [\langle z, Qz \rangle_H + \langle K_i z + \bar{u}, R(K_i z + \bar{u}) \rangle_U] d\sigma - \langle z(T), Gz(T) \rangle_H .$$

Substituting from theorem I gives

$$\langle z(t),P_i(t)z(t)\rangle_H - \langle z(t),P_{i+1}(t)z(t)\rangle_H = \langle z(T),Gz(T)\rangle_H +$$
$$\int_t^T\left[\langle z,Qz\rangle_H - 2\langle z,P_iB\bar{u}\rangle_H\right]d\sigma - \int_t^T\left[\langle z,Qz\rangle_H + \langle z,K_i^*RK_iz\rangle_H\right.$$
$$\left. + 2\langle\bar{u},RK_iz\rangle_U + \langle\bar{u},R\bar{u}\rangle_U\right]d\sigma - \langle z(T),Gz(T)\rangle_H$$
$$= \int_t^T\left[\langle\bar{u},R\bar{u}\rangle_U - 2\langle\bar{u},(B^*P_i + RK_i)z\rangle_U\right]d\sigma,$$

here the fact that $Q_i = Q + K_i^*RK_i$ has been used. So, if K_i is chosen equal to $-R^{-1}B^*P_i$ the second term under the integral becomes zero with the result that

$$\langle z(t),P_i(t)z(t)\rangle_H - \langle z(t),P_{i+1}(t)z(t)\rangle_H = \int_t^T\langle\bar{u},R\bar{u}\rangle_U\,d\sigma.$$

Since R is positive definite this must be greater than zero unless $u(\sigma) = 0$ for all σ, which will later be shown to correspond to the optimal control case. Therefore

$$\langle z(t),P_{i+1}(t)z(t)\rangle_H \leq \langle z(t),P_i(t)z(t)\rangle_H$$

or, putting $t = 0$ and using (a)

$$J(u_{i+1}) \leq J(u_i) .$$

Having proved that it is possible to construct a sequence of controls of decreasing costs it is necessary to prove next that both the cost and the control converge to some limit as i is increased indefinitely. This is done in the following theorem.

Theorem III

$P_\infty(t)$ exists and $P_i(t) \to P_\infty(t)$ in the strong sense as $i \to \infty$ moreover, $P_\infty(t)$ satisfies

$$\langle z(t),P_\infty(t)z(t)\rangle_H = \int_t^T\left[\langle z,Qz\rangle_H + \langle u,Ru\rangle_U - \langle\bar{u},R\bar{u}\rangle_U\right]d\sigma$$
$$+ \langle z(T),Gz(T)\rangle_H$$

and the control

$$u_\infty = -R^{-1}B^*P_\infty z$$

is the optimal control.

Proof:

$$\langle z(t),P_0(t)z(t)\rangle_H = \langle z(T),Gz(T)\rangle_H + \int_t^T\langle z,Qz\rangle_H\,d\sigma,$$

this coming from theorem I when $\bar{u} = 0$, $i = 0$, and $K_0 = 0$. Q and G are bounded, A is the infinitesimal generator of a strongly continuous semigroup, so $P_0(t)$ is uniformly bounded. Therefore $P_i(t)$ is a non-increasing sequence of self-adjoint operators, and is also uniformly bounded above and below; hence $P_i(t)$ converges strongly to the self-adjoint operator $P_\infty(t)$.

We must show how the result of theorem III implies the existence of an optimal feedback controller $u_\infty = K_\infty z$ which gives rise to a cost less than or equal to that of any other control. In theorem II we set

$$K_i = -R^{-1}B^*P_{i-1},$$

hence

$$K_i + R^{-1}B^*P_\infty = R^{-1}B^*(P_\infty - P_{i-1}) \quad .$$

Now, K_i and P_i are uniformly bounded and P_i converges strongly to P_∞, therefore $K_i \to -R^{-1}B^*P_\infty$ strongly. By definition

$$Q_i = Q + K_i^*RK_i = Q + P_{i-1}BR^{-1}B^*P_{i-1}.$$

B and R^{-1} are uniformly bounded so Q_i converges strongly to

$$Q_\infty = Q + P_\infty BR^{-1}B^*P_\infty \quad .$$

The strong convergence of P_i , K_i and Q_i can now be used in theorem I together with the Lesbegue dominated convergence theorem to yield

$$\langle z(t),P_\infty(t)z(t)\rangle_H = \langle z(T),Gz(T)\rangle_H + \int_t^T \left[\langle z,(Q + P_\infty BR^{-1}B^*P_\infty)z\rangle_H \right.$$
$$\left. -2\langle z,P_\infty B\bar{u}\rangle_H\right] d\sigma .$$

As $i \to \infty$ the perturbed control

$$u = K_\infty z + \bar{u} = -R^{-1}B^*P_\infty z + \bar{u} ,$$

hence, on substituting for $B^*P_\infty z$ on the right hand side,

$$\langle z(t),P_\infty(t)z(t)\rangle_H = \langle z(T),Gz(T)\rangle_H + \int_t^T \left[\langle z,Qz\rangle_H + \langle \bar{u}-u,R(\bar{u}-u)\rangle_U \right.$$
$$\left. - 2\langle \bar{u}-u,R\bar{u}\rangle_U\right] d\sigma .$$

Putting $t = 0$ and noting that from theorem I with $t = 0$ the left hand side $= J(u_\infty)$ one obtains

$$J(u_\infty) = \langle z(T), Gz(T)\rangle_H + \int_0^T \left[\langle z, Qz\rangle_H + \langle u, Ru\rangle_U\right] d\sigma - \int_0^T \langle \bar{u}, R\bar{u}\rangle_U \, d\sigma .$$

Therefore, from the definition of $J(u)$, (1.1.8),

$$J(u) - J(u_\infty) = \int_0^T \langle \bar{u}, R\bar{u}\rangle_U \, d\sigma .$$

R is a positive definite operator so the right hand side is non-negative and

$$J(u_\infty) \leq J(u)$$

thus proving that

$$u_\infty = K_\infty z = - R^{-1} B^* P_\infty z$$

is the optimal control.

It should be noted that P_∞ may be replaced by P_i in the above proof provided that $K_i = - R^{-1} B^* P_i$. As a result $u_i = K_i z = u_\infty$ which means that in generating this series of controls it may be possible to reach the optimal in a finite number of steps.

Section 2. The infinite dimensional Riccati equation.

The optimal control for a finite dimensional system can be found by solving the matrix Riccati equation, Athans & Falb(1966). Therefore it would be of interest to see how the optimal control derived in section 1 corresponds to this result. The full proof is given by Curtain & Pritchard(1975) but here we shall give an account of how a weak inner product version of the infinite dimensional Riccati equation can be derived.

Consider the inner product

$$\langle x, P_i(t) y\rangle_H \qquad (1.2.1)$$

where x and y are arbitrary elements of $\mathbf{D}[A]$. Now, as $i \to \infty$, we have proved that $P_i(t)$ converges strongly to $P_\infty(t)$, so using

(1.1.12) as $i \to \infty$, the form (1.2.1) becomes

$$\langle x, P_\infty(t)y\rangle_H = \left\langle x, \left[S^*_\infty(T,t)GS_\infty(T,t) + \int_t^T S^*_\infty(\sigma,t)Q_\infty(\sigma)\, S_\infty(\sigma,t)d\sigma \right] y \right\rangle_H .$$

This equation can be differentiated with respect to t to obtain

$$\langle x, \dot{P}_\infty(t)y\rangle_H = \left\langle x, \left\{ \frac{\partial S^*_\infty}{\partial t}(T,t)GS_\infty(T,t) + S^*_\infty(T,t)G\frac{\partial S_\infty}{\partial t}(T,t) \right.\right.$$
$$- S^*_\infty(t,t)Q_\infty(t)S_\infty(t,t) + \int_t^T \left[\frac{\partial S^*_\infty}{\partial t}(\sigma,t)Q_\infty(\sigma)S_\infty(\sigma,t) \right.$$
$$\left.\left. + S^*_\infty(\sigma,t)Q_\infty(\sigma)\frac{\partial S_\infty}{\partial t}(\sigma,t)\right] d\sigma \right\} y \Big\rangle_H ,$$

see Curtain and Pritchard(1975). Using the relationships (1.1.7) we get

$$\langle x, \dot{P}_\infty(t)y\rangle_H = \left\langle x, \left\{ -A^*_\infty(t)S^*_\infty(T,t)GS_\infty(T,t) - S^*_\infty(T,t)GS_\infty(T,t)A_\infty(t) \right.\right.$$
$$- Q_\infty(t) - \int_t^T \left[A^*_\infty(t)S^*_\infty(\sigma,t)Q_\infty(\sigma)S_\infty(\sigma,t) \right.$$
$$\left.\left. + S^*_\infty(\sigma,t)Q_\infty(\sigma)S_\infty(\sigma,t)A_\infty(t)\right] d\sigma \right\} y \Big\rangle_H$$
$$= \left\langle x, \left[-A^*_\infty(t)P_\infty(t) - P_\infty(t)A_\infty(t) - Q_\infty(t) \right] y \right\rangle_H .$$

Now,

$$Q_\infty(t) = Q + K^*_\infty(t)RK_\infty(t) = Q + P_\infty(t)\, B\, R^{-1}\, B^*\, P_\infty(t)$$

and

$$A_\infty(t) = A - BR^{-1}\, B^*\, P_\infty(t) .$$

Therefore

$$\left\langle x, \left[\dot{P}_\infty(t) + P_\infty(t)A + A^*P_\infty(t) - P_\infty(t)BR^{-1}B^*P_\infty(t) + Q \right] y \right\rangle_H = 0 . \qquad (1.2.2)$$

We also have the end condition on $P_\infty(t)$

$$P_\infty(T) = G . \qquad (1.2.3)$$

(1.2.2) and (1.2.3) together form the inner product version of the infinite dimensional Riccati equation which corresponds to the finite dimensional result. If $P_\infty(t)$ can be differentiated directly there is no need to use the inner product form and one consequently obtains a strong version of the Riccati equation.

Section 3. Control on the infinite interval.

In the derivation of the optimal control presented in section 1, only the finite interval was considered in the cost function. In order to extend these results to include control on the infinite interval it is necessary to introduce some additional assumptions. Firstly the cost function is redefined as

$$J^{\infty}(u) = \int_0^{\infty} \left[\langle z(t), Qz(t) \rangle_H + \langle u(t), Ru(t) \rangle_U \right] dt \qquad (1.3.1)$$

where Q and R are now time invariant. One must also assume that the systemis optimisable relative to Q, Lukes & Russell(1969). The system is said to be optimisable relative to Q if there exist $k_c \geq 0$ and a bounded operator K_c^{∞}, independent of t such that the solution z(t)

$$z(t) = T_t^c z_0 + \int_0^t T_{t-\sigma}^c B K_c^{\infty} z(\sigma) d\sigma \qquad (1.3.2)$$

corresponding to the control

$$u_c(t) = K_c^{\infty} z(t)$$

gives values of $J^{\infty}(u_c)$ defined by (1.3.1) satisfying

$$J^{\infty}(u_c) \leq k_c \| z_0 \|^2$$

for all z_0 in D(A). In these expressions T_t^c is the strongly continuous semigroup generated by the operator

$$A_c = A + BK_c^{\infty}.$$

It will be shown that there exists a sequence of controls u_k^{∞}, k=0,1,2... and bounded operators P_k^{∞} such that

$$J^{\infty}(u_k^{\infty}) = \langle z_0, P_k^{\infty} z_0 \rangle_H$$

and

$$J^{\infty}(u_{k+1}^{\infty}) \leq J^{\infty}(u_k^{\infty}) \quad .$$

Firstly it will be proved that if

$$\langle x, P_o^{t_1}(t)y\rangle_H = \langle x, \int_t^{t_1} T_{\sigma-t}^* Q_o T_{\sigma-t}\, d\sigma\, y\rangle_H, \qquad (1.3.3)$$

where $x,y \in H$ and

$$Q_o = K_o^{\infty *} R K_o^{\infty} + Q ,$$

then $P_o^{\infty}(t)$ exists, $P_o^{\infty}(t) = P_o^{\infty}$ is independent of t and

$$\langle z_o, P_o^{t_1}(0) z_o\rangle_H \leq \langle z_o, P_o^{\infty} z_o\rangle \leq k_o \|z_o\|^2 .$$

From (1.3.3)

$$\langle x, P_o^{t_1}(t+\alpha)y\rangle_H = \langle x, \int_{t+\alpha}^{t_1} T_{s-t-\alpha}^{o *} Q_o T_{s-t-\alpha}^{o}\, ds\, .\, y\rangle_H$$
$$= \langle x, \int_t^{t_1-\alpha} T_{\sigma-t}^{o *} Q_o T_{\sigma-t}\, d\sigma\, .\, y\rangle_H$$

where $\sigma = s - \alpha$,

$$= \langle x, P_o^{t_1-\alpha}(t) y\rangle_H \qquad (1.3.4)$$

since Q_o is independent of t.

$$u_o^{\infty}(t) = K_o^{\infty} z(t) ,$$

so, from (1.3.3),

$$\int_t^{t_1} \left[\langle z(s), Qz(s)\rangle_H + \langle u_o^{\infty}(s), Ru_o^{\infty}(s)\rangle_U\right] ds = \int_t^{t_1} \langle z(s), Q_o z(s)\rangle_H ds$$
$$= \langle z(t), P_o^{t_1}(t) z(t)\rangle_H . \qquad (1.3.5)$$

Since Q_o is independent of t and is a positive semi-definite operator, the integrand in (1.3.5) is non-negative, so $\langle z(t), P_o^{t_1}(t)z(t)\rangle_H$ is a non-decreasing function of t_1 and a non-increasing function of t. Hence, using the fact that the system is optimisable relative to Q,

$$\langle z(t), P_o^{t_1}(0) z(t)\rangle_H \leq \langle z_o, P_o^{\infty}(0) z_o\rangle_H = J^{\infty}(u_o^{\infty}) \leq k_o \|z_o\|^2$$

for $t_1 \geq t$, $z_o \in D(A)$. Therefore as $t_1 \to \infty$, $P_o^{t_1}(t)$ is non-decreasing and is bounded, so $P_o^{t_1}(t) \to P_o^{\infty}(t)$ strongly as $t_1 \to \infty$. Now, using (1.3.4),

$$P_o^{\infty}(t) = \lim_{t_1 \to \infty} P_o^{t_1}(t) = \lim_{t_1 \to \infty} P_o^{t_1 - t}(0) = P_o^{\infty}(0).$$

Hence it has been shown that the limit $P_o^{\infty}(t)$ exists and is independent

of t.

The next step is to extend this result for a series of controllers u_i^∞ and to show that

$$J^\infty(u_{i+1}^\infty) \leq J^\infty(u_i^\infty) .$$

Consider a control

$$u_i^\infty(t) = K_i^\infty z(t),$$

so the cost associated with this control is given by

$$J^\infty(u_i^\infty) = \int_0^\infty \left[\langle z(t), Qz(t)\rangle_H + \langle u_i^\infty(t), Ru_i^\infty(t)\rangle_U\right]dt$$
$$= \langle z_0, P_i^\infty z_0\rangle_H . \qquad (1.3.6)$$

The existence of P_i^∞ is proved in exactly the same way as that of P_0^∞ since K_i^∞ is independent of t. Now consider a perturbation of the contrd such that

$$u_{i+1}(t) = K_i^\infty z(t) + \bar{u}(t).$$

Therefore

$$J^{t_1}(u_{i+1}) = \int_0^{t_1}\left[\langle z(t), Qz(t)\rangle_H + \langle K_i^\infty z(t) + \bar{u}(t), R(K_i^\infty z(t) + \bar{u}(t))\rangle_U\right]dt$$
$$= \int_0^{t_1}\left[\langle z(t), Q_i z(t)\rangle_H + 2\langle \bar{u}(t), RK_i^\infty z(t)\rangle_H + \langle \bar{u}(t), R\bar{u}(t)\rangle_U\right]dt \qquad (1.3.7)$$

However, putting t = 0 in theorem I yields

$$\langle z_0, P_i^{t_1}(0)z\rangle_H = \int_0^{t_1}\left[\langle z(t), Q_i z(t)\rangle_H - 2\langle z(t), P_i^{t_1}(t)B\bar{u}(t)\rangle_H\right]dt.$$

Using this result to substitute for $\int_0^{t_1}\langle z(t), Q_i z(t)\rangle dt$ in (1.3.7)

$$J^{t_1}(u_{i+1}) = \langle z_0, P_i^{t_1}(0)z_0\rangle_H + \int_0^{t_1} 2\langle z(t), P_i^{t_1}(t)B\bar{u}(t)\rangle_H + 2\langle \bar{u}(t), RK_i^\infty z(t)\rangle_U$$
$$+ \langle \bar{u}(t), R\bar{u}(t)\rangle_U\Big] dt.$$

However, from (1.3.6), $J^\infty(u_i^\infty) = \langle z_0, P_i^\infty z\rangle_H$, so

$$J^{t_1}(u_{i+1}) - J^\infty(u_i^\infty) = \langle z_0, (P_i^{t_1}(0) - P_i^\infty)z_0\rangle_H$$
$$+ \int_0^{t_1}\left[\langle \bar{u}(t), R\bar{u}(t)\rangle_U + 2\langle \bar{u}(t), (B^*P_i^{t_1}(t) + RK_i^\infty)z(t)\rangle_U\right] dt$$

$$= \langle z_0, (P_i^{t_1}(0) - P_i^\infty)z_0\rangle_H$$
$$+ \int_0^{t_1}\Big[\langle \bar{u}(t) + R^{-1}(B^*P_i^{t_1}(t) + RK_i^\infty)z(t), R\left[\bar{u}(t) + R^{-1}(B^*P_i^{t_1}(t) + RK_i^\infty)z(t)\right]\rangle_U$$
$$- \langle (B^*P_i^{t_1}(t) + RK_i^\infty)z(t), R^{-1}(B^*P_i^{t_1}(t) + RK_i^\infty)z(t)\rangle_U\Big] dt.$$

If we now choose

$$\bar{u}(t) = - R^{-1}(B^*P^{t_1}(t) + RK^{\infty})z(t) \qquad —$$

we obtain the result

$$J^{t_1}(u_{i+1}) - J^{\infty}(u_i^{\infty}) = \langle z_0, (P_i^{t_1}(0) - P_i^{\infty})z_0 \rangle_H - \int_0^{t_1} \langle v(t), R^{-1}v(t) \rangle \, dt \qquad (1.3.8)$$

where

$$v(t) = (B^*P_i^{t_1}(t) + RK_i^{\infty})z(t) \quad .$$

Let $t_1 \to \infty$, it is known that $P_i^{t_1}(0) \to P_i^{\infty}$ strongly, so the first term on the right hand side of (1.3.8) tends to zero.

$$u_{i+1}(t) = K_i^{\infty}z(t) + \bar{u}(t)$$

$$= K_i^{\infty}z(t) - R^{-1}B^*P_i^{t_1}(t)z(t) - K_i^{\infty}z(t) = - R^{-1}B^*P_i^{t_1}(t)z(t) \ ,$$

R^{-1} and B^* are bounded operators so $u_{i+1}(t)$ converges strongly to $- R^{-1}B^*P_i^{\infty}z(t) = u_{i+1}^{\infty}(t)$. Therefore, as a result of letting $t_1 \to \infty$

$$J^{\infty}(u_{i+1}^{\infty}) - J^{\infty}(u_i^{\infty}) = -\int_0^{\infty} \langle v(t), R^{-1}v(t) \rangle \, dt.$$

R^{-1} is positive definite so

$$J^{\infty}(u_{i+1}^{\infty}) \leq J^{\infty}(u_i) \ ,$$

the equality only holding if $v(t) = 0$ for all $t \in [0, \infty)$. $v(t)$ is identically zero only if

$$B^*P_i^{\infty} + RK_i^{\infty} = 0 \ , \text{ or } K_i^{\infty} = - R^{-1}B^*P_i^{\infty} \ .$$

In this case it can be shown that P_i^{∞} satisfies the Riccati equation and that the optimal has been reached in a finite number of steps.

Hence, if one can choose a control $u_0^{\infty}(t) = K_0^{\infty}z(t)$ such that $J^{\infty}(u_0^{\infty}) \leq k_0 \| z_0 \|^2$ it is then possible to generate a series of controls u_i^{∞} such that $J^{\infty}(u_{i+1}) \leq J^{\infty}(u_i)$.

$$J^{\infty}(u_0^{\infty}) = \langle z_0, P_0^{\infty}z_0 \rangle_H$$

so P_i^{∞} is a non-increasing sequence of operators, bounded below by zero as it is impossible to have a negative cost, thence P_i^{∞} converges strongly to some limit P_{∞}^{∞}. It is now necessary to show

that this limit yields the optimal control

$$u_{\infty}^{\infty}(t) = - R^{-1}B^{*}P_{\infty}^{\infty}z(t)$$

such that

$$J^{\infty}(u_{\infty}^{\infty}) \leq J^{\infty}(u)$$

for all $u \in U$.

Using the Lebesgue dominated convergence theorem in conjunction with theorem I, the following result is obtained

$$\langle z(t),P_{\infty}^{t_1}(t)z(t)\rangle_H = \int_t^{t_1}\left[\langle z(s),Q_{\infty}^{t_1}(s)z(s)\rangle_H - 2\langle z(s),P_{\infty}^{t_1}(s)B\bar{u}(s)\rangle_U\right] ds \tag{1.3.9}$$

where the control

$$u(t) = - R^{-1}B^{*}P_{\infty}^{t_1}(t)z(t) + \bar{u}(t) = K_{\infty}^{t_1}(t)z(t) + \bar{u}(t)$$

and

$$Q_{\infty}^{t_1}(t) = Q + P_{\infty}^{t_1}(t)BR^{-1}B^{*}P_{\infty}^{t_1}(t).$$

We now substitute this control into the cost function

$$J^{t_1}(u) = \int_0^{t_1}\left[\langle z(t),Q_{\infty}^{t_1}(t)z(t)\rangle_H + 2\langle\bar{u}(t),RK_{\infty}^{t_1}(t)z(t)\rangle_U + \langle\bar{u}(t),R\bar{u}(t)\rangle_U\right]dt \quad . \tag{1.3.10}$$

Setting $t = 0$ in (1.3.9) gives

$$\langle z_0,P_{\infty}^{t_1}(0)\ z_0\rangle_H = \int_0^{t_1}\left[\langle z(t),Q_{\infty}^{t_1}(t)z(t)\rangle_H + 2\langle z(t),RK_{\infty}^{t_1}(t)z(t)\rangle_H\right]dt \tag{1.3.11}$$

where the relationship $K_{\infty}^{t_1}(t) = - R^{-1}P_{\infty}^{t_1}(t)B$ has been used.

Finally, subtracting (1.3.11) from (1.3.10) yields

$$J^{\infty}(u) - \langle z_0,P_{\infty}^{t_1}(0)z_0\rangle_H = \int_0^{t_1}\langle\bar{u}(t),R\bar{u}(t)\rangle_U dt \geq 0 \tag{1.3.12}$$

as R is positive definite. If we let $t_1 \to \infty$, $P_{\infty}^{t_1}(0)$ converges strongly to P_{∞}^{∞}, $u(t)$ converges strongly to $- R^{-1}B^{*}P_{\infty}^{\infty}z(t)$ since R and B^{*} are bounded, so, as long as $J^{\infty}(u) < \infty$,

$$J^{\infty}(u) \geq J^{\infty}(u_{\infty}^{\infty})$$

this following from (1.3.12) where

$$\langle z_0,P_{\infty}^{\infty}z_0\rangle_H = J^{\infty}(u_{\infty}^{\infty}).$$

Therefore

$$u_{\infty}^{\infty}(t) = - R^{-1}B^{*}P_{\infty}^{\infty}z(t)$$

is the optimal control for the infinite interval problem.

Section 4. Conclusions.

It has been proved, therefore, that under suitable assumptions an optimal control exists for both the finite and infinite interval and can be found by constructing a series of controls that converges strongly to the optimal. If one is justified in using the strong, instead of the mild, forms the optimal control is given by the standard Riccati equation that is derived by dynamic programming, Wang(1964). Moreover, this method of generating a sequence of controls is useful from a computational point of view even when the strong form of the Riccati equation is valid. For example the case of a finite dimensional system considered over the infinite interval results in having to solve a quadratic matrix equation in P with the condition that P must be positive definite; this calculation can be difficult and time consuming. The solution by the means presented in this chapter gives a numerical method that is guaranteed to converge involving only a linear equation in P to be solved at every iteration. It is interesting to note that the sequence is identical to that generated if one attempts to solve the matrix Riccati equation directly using Newton's method, with, of course, the advantage of being certain that the iteration converges to the correct root.

The main drawback of the optimal control is that the feedback law $u_\infty = K_\infty z = -R^{-1}B^*P_\infty z$ can only be implemented if one knows $z(t)$ completely for all $t \in [0,T]$. This may be difficult for finite dimensional systems but will be impossible for those of infinite dimension except in special circumstances where it is possible to build a distributed sensor, an example of which is mentioned in chapter 2. The problems of feedback control with limited knowledge of the state are considered in detail in subsequent chapters.

CHAPTER 2

THE CONSTRAINED OPTIMAL CONTROL AND BOUNDS ON THE COST FUNCTION

Section 1. Introduction.

In chapter 1 it was shown how the optimal control for the linear quadratic problem can be derived for both finite and infinite dimensional systems. However, there are many practical difficulties that can arise either from the particular physical configuration of the system or from the quantity and complexity of the numerical calculations necessary to derive the optimal control. One serious difficulty in implementing the optimal control is due to the fact that it is necessary to have feedback of all the state variables; this may pose considerable problems in finite dimensional systems but it will be insuperable in nearly all those of infinite dimension. One can also be faced with an obstacle to applying the control action; to be practically feasible it will have to be applied at the boundary of a distributed parameter system, not throughout the space occupied by the system. For example, the control of a vibrating medium governed by the wave equation requires the measurement of displacement and velocity at all points within the medium. Similarly the optimal control of one dimensional heat flow in a bar necessitates sensing the temperature at all points along the bar. However, this is a case where it would be feasible to build the optimal controller provided the control action takes place at the boundaries. Here one needs the convolution of some function with temperature along the bar, Pritchard & Mayhew(1970), and this could be achieved by using a strip of suitable material of varying thickness or width which is fixed to the bar from one end to the other. These sorts of sys-

tem, though, constitute special cases which can only be treated on their own merits, in general one must accept that the optimal control will not be realisable. Some specific problems of discrete sensors in distributed parameter systems will be considered later.

Even in systems which can be described by a finite number of ordinary differential equations it is quite likely that all the state variables cannot be measured, for example in the control of multi-stage chemical processes or aeroplane dynamics. Although some measurements are feasible it might well be uneconomic to modify the plant in order to introduce sensors unless their presence would result in very great benefits. In this case some prior analysis has to be carried out to evaluate the disadvantages of omitting the sensing of certain state variables. Systems governed by partial differential equations can often be approximated by a finite number of ordinary differential equations and indeed this is frequently the only way to tackle the computational problems. Hence, in many situations one is confronted with the problem of constructing an optimal feedback controller for a system governed by a finite number of ordinary differential equations when one only has limited knowledge of the state of the system.

One way of approaching this problem is by the construction of a dynamic observer, Luenberger(1966). It can be shown that if one is observing an n^{th} order system with m outputs it is possible, by means of a $(n-m)^{th}$ order dynamic observer, to obtain a signal that converges asymptotically to the state of the original system. Even though the observer can be made to react with arbitrarily small time constants, the use of this method to obtain feedback from the estimated state must increase the cost function by a finite amount com-

pared with its optimal value, Bongiorno & Youla(1968,1970). However, these authors also show that one can make the value of the cost function arbitrarily close to its optimal by building an n^{th} order observer but the gains involved take on very high values. There are two main disadvantages to such a technique. Firstly, if one is not able to measure all the state variables of the n^{th} order system, it might well not be easy to obtain access to all n state variables of the observer. Obviously if $n \to \infty$ the difficulties of knowing the state of a distributed parameter observer are just as great as those of the original system. Secondly, the necessity for very high gains will cause considerable problems in practice, mainly that the amplifiers are likely to saturate so that the system is no longer linear, thus invalidating the optimal control.

An important question that must be asked is, in what sense does an optimal control exist when one has incomplete knowledge of the state? First a cost function has to be defined and we shall consider one which takes the standard quadratic form. This has the advantage of allowing direct comparison with the result arising from the case when there is complete observation of the state. We shall now investigate this question in detail in the next section.

Section 2. The constrained optimal control.

It is known from chapter 1 that having access to all the state variables gives an optimal control that yields the global minimum of the cost function. We shall derive certain sufficient conditions under which the partially observed system has a meaningful optimal feedback control that also gives a global minimum for the cost function and is of a form that is practically realisable. Only finite dimensional systems will be considered here as it is

theoretically considerably more manageable and also because, as stated earlier, almost any computation carried out in the analysis of infinite dimensional systems will involve some finite approximation.

Consider the general linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$
$$x(0) = x_0 \qquad (2.2.1)$$

where $x(t)$ is an $n\times 1$ state vector, $u(t)$ is an $r\times 1$ control vector and $A(t)$, $B(t)$ are matrices of the appropriate dimensions. The limitation on the knowledge of the state is represented by assuming that there are only m independent outputs where $0<m<n$, that is

$$y(t) = C(t)x(t) \qquad (2.2.2)$$

where $y(t)$ is an $m\times 1$ output vector and $C(t)$ is an $m\times n$ matrix. Finally, since we are only considering linear systems, the control $u(t)$ is taken to be a linear function of the output, $y(t)$,

$$u(t) = K(t)y(t) \qquad (2.2.3)$$

where $K(t)$ is an $r\times m$ matrix of feedback gains. The cost fuction J, for reasons stated earlier, is taken to be of the quadratic form

$$J = x'(T)Gx(T) + \int_0^T \left[x'(t)Q(t)x(t) + u'(t)R(t)u(t)\right] dt, \qquad (2.2.4)$$

G and $Q(t)$ are $n\times n$ symmetric positive semi-definite matrices and $R(t)$ is an $r\times r$ symmetric positive definite matrix. It is well known, Athans & Falb(1966), that the optimal control for the totally observed system is

$$u(t) = -R^{-1}(t)\, B'(t)\, P(t)\, x(t) \qquad (2.2.5)$$

where $P(t)$ is the symmetric positive definite matrix satisfying the Riccati equation

$$\dot{P}(t)+P(t)A(t)+A'(t)P(t)-P(t)B'(t)R^{-1}(t)B(t)P(t)+Q(t)=0$$
$$P(T)=G\ . \qquad (2.2.6)$$

Now, (2.2.1), (2.2.5) and (2.2.6) define optimal trajectories

$$x(t)=x^*(t),\ y(t)=y^*(t),\ u(t)=u^*(t)$$

for $t\in[0,T]$. If it is possible to find a K(t) such that

$$u^*(t) = K(t)y^*(t) = K(t)C(t)x^*(t) \qquad (2.2.7)$$

then the optimal controller can be constructed. In general this can be done if K(t) is allowed to vary with time because (2.2.7) requires r conditions to be met while K(t) contains $m\times r \geq r$ elements that can be varied. However, the solution of the linear equation for K(t) can become impossible if, for example, $y^*(t)=0$ when $u^*(t)\neq 0$. In this case some of the elements of K(t) will tend to infinity which is not feasible in practice. If some upper bounds are put on the absolute values of the elements of K(t) one would be left with a different sort of optimisation problem which is best dealt with by use of Pontryagin's maximum principle, Pontryagin et al(1962). However, letting K(t) vary with time like this would not be particularly realistic in a practical situation as it would involve open loop control of the elements of K(t). This negates the advantages of having a feedback controller, it would be better to use open loop control on u(t) directly and use a computer on line to feed $u^*(t)$ into the system. Hence we shall consider the problem of finding the time invariant K that minimises J, this retains the advantages of a feedback controller which can still perform reasonably well in changing conditions. This will coincide with the optimal control if the system is time invariant and considered over the infinite interval provided that m=n and C is non-singular. In this case the optimal control is known to be a feedback law that is independent of time, Athans & Falb(1966). If C can be inverted it is easy to reconstruct the state, $x=C^{-1}y$, and thence apply the optimal control. However,

this work will be restricted to the case where m<n.

We shall now define what is meant by an optimal control when the knowledge of the state is limited and the feedback matrix is constant over time. There will be said to be a realisable constrained optimal control if there exists a time invariant $K=K^*$ with finite norm such that

$$\inf_{K\in\mathcal{K}} \left[J(K,x_o,T) \right] = J^*(K^*,x_o,T) \qquad (2.2.8)$$

where $\mathcal{K}$ is the set of real r×m matrices. It should be noted that this control K^* will depend on the initial state x_o and this is a fundamental disadvantage of partially observed systems compared with those that are totally observed.

There is no immediate reason why letting $\|K\|\to\infty$ should not lead to a finite value of J which, in turn, might be the optimal value J^*. However, if it can be shown that there exists a $K\in\mathcal{K}$ such that $J(K,x_o,T)$ is finite and that

$$\lim_{\|K\|\to\infty} J(K,x_o,T) = \infty$$

then there will be a realisable constrained optimal control. The contribution to J due directly to the control u(t), J_u, is given by

$$J_u = \int_o^T u'(t)R(t)u(t)dt = \int_o^T x'(t)C'(t)K'R(t)KC(t)x(t)dt \qquad (2.2.9)$$

and as $\|K\|\to\infty$ it could be possible for $x(t)\to 0$ in such a way that this integral is finite. The other terms in (2.2.4) contributing towards J do not contain K explicitly and if $x(t)\to 0$ as $\|K\|\to\infty$ they will tend to zero too, this following from the fact that the norm of x(t) is bounded above by a negative exponential as a result of (1.1.7). Therefore one of the main objectives must be to find under what conditions

$$\lim_{\|K\|\to\infty} J_u = \infty .$$

If the system is allowed to be time varying it is very difficult to say much about the behaviour of the integrand in (2.2.9). At any time, t, there is a subspace of $\mathbb{R}^n$, $S_o(t)$, for which $x(t) \in S_o(t)$ implies that $u(t)=KC(t)x(t)=0$. If $x(t) \notin S_o(t)$ it is possible in the time variant case for x to become an element of S_o in finite time and remain so thereafter. This makes the problem of finding bounds on J_u more difficult, but if an upper limit is placed on the rate at which C(t) can vary it is possible to show that J_u has a finite lower bound as $\|K\| \to \infty$, but not that $J_u \to \infty$. Hence we shall restrict ourselves to the case of time invariant systems since the trajectory of x(t) can then be written explicitly. Also this type of problem is more common in practice. Usually the period of response is smaller than the time scale on which the the parameters of the system vary. Moreover, systems in which the time variations are known accurately are rare; it is more likely that the mean values are known with some random fluctuations superimposed and this is a problem needing a different kind of approach.

(2.2.1), (2.2.2) and (2.2.3) may be combined to give

$$\begin{aligned} \dot{x}(t) &= (A+BKC)x(t) = Fx(t) \\ x(0) &= x_o. \end{aligned} \qquad (2.2.10)$$

The behaviour of x(t) will depend on the eigenvalues of F and whether they are distinct or if the characteristic equation of F has repeated roots. The effect of letting $\|K\| \to \infty$ will be allowed for by replacing K with αK, where α is a positive real scalar, and then letting $\alpha \to \infty$. This is valid because the only way a matrix of finite dimensions can have an infinite norm is for at least one of its elements to be infinite. If the eigenvalues of BKC are μ_i then, as $\alpha \to \infty$, the eigenvalues of F tend to $\alpha(\mu_i + \epsilon_i)$ where ϵ_i is of the order $1/\alpha$. If

If the system is allowed to be time varying it is very difficult to say much about the behaviour of the integrand in (2.2.9). At any time, t, there is a subspace of $\mathbb{R}^1$, $S_o(t)$, for which $x(t) \in S_o(t)$ implies that $u(t)=KC(t)x(t)=0$. If $x(t) \notin S_o(t)$ it is possible in the time variant case for x to become an element of S_o in finite time and remain so thereafter. This makes the problem of finding bounds on J_u more difficult, but if an upper limit is placed on the rate at which C(t) can vary it is possible to show that J_u has a finite lower bound as $\|K\| \to \infty$, but not that $J_u \to \infty$. Hence we shall restrict ourselves to the case of time invariant systems since the trajectory of x(t) can then be written explicitly. Also this type of problem is more common in practice. Usually the period of response is smaller than the time scale on which the the parameters of the system vary. Moreover, systems in which the time variations are known accurately are rare; it is more likely that the mean values are known with some random fluctuations superimposed and this is a problem needing a different kind of approach.

(2.2.1), (2.2.2) and (2.2.3) may be combined to give

$$\begin{aligned} \dot{x}(t) &= (A + BKC)x(t) = Fx(t) \\ x(0) &= x_o. \end{aligned} \qquad (2.2.10)$$

The behaviour of x(t) will depend on the eigenvalues of F and whether they are distinct or if the characteristic equation of F has repeated roots. The effect of letting $\|K\| \to \infty$ will be allowed for by replacing K with αK, where α is a positive real scalar, and then letting $\alpha \to \infty$. This is valid because the only way a matrix of finite dimensions can have an infinite norm is for at least one of its elements to be infinite. If the eigenvalues of BKC are μ_i then, as $\alpha \to \infty$, the eigenvalues of F tend to $\alpha(\mu_i + \epsilon_i)$ where ϵ_i is of the order $1/\alpha$. If

any of these have positive real parts and x_0 has a component in the direction of the corresponding eigenvector then, as $\alpha \to \infty$, $\|x(t)\| \to \infty$ as will J_ω. The interesting case occurs when the only eigenvalues that enter the expression for u(t) have negative real parts. The integrand in (2.2.9) is

$$\alpha^2 x'(t)C'K'RKCx(t)$$

and as $\alpha \to \infty$ $\|x(t)\| \to 0$ and $\alpha^2 \to \infty$. Hence we must look more closely at the expression for x(t) to find the limit of J_ω.

The solution for x(t) is given in Ogata(1967) and the i^{th} component of the state will be given by an expression of the form

$$x_i(t) = \sum_{\ell=1}^{n_c} \sum_{k=0}^{m_\ell - 1} a_{i\ell k} t^k e^{\lambda_\ell t} \qquad (2.2.11)$$

where F has n_c distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n_c}$ and λ_ℓ has multiplicity m_ℓ. The elements of u(t)= KCx(t) are linear combinations of the state variables, hence we may write

$$u_i(t) = \alpha \sum_{\ell=1}^{n_c} \sum_{k=0}^{m_\ell - 1} b_{i\ell k} t^k e^{\lambda_\ell t} \quad . \qquad (2.2.12)$$

$$J_\omega = \int_0^T u'(t)Ru(t)dt$$

so, from (2.2.12)

$$J_\omega = \int_0^T \alpha^2 \sum_{i,j=1}^{n} r_{ij} \sum_{\ell=1}^{n_c} \sum_{k=0}^{m_\ell - 1} \sum_{\ell'=1}^{n_c} \sum_{k'=0}^{m_{\ell'} - 1} b_{i\ell k} b_{j\ell' k'} t^{(k+k')} e^{(\lambda_\ell + \lambda_{\ell'})t} dt \quad .$$

However, we are going to consider the case where $\alpha \to \infty$, that is $\lambda_\ell \to \alpha(\mu_\ell + \zeta_\ell)$. m<n so BKC is not of full rank hence some of the μ_ℓ will be zero, that is $\lambda_\ell = \alpha\zeta_\ell$. As a result, some of the λ_ℓ's will not have real parts that tend to $-\infty$ as $\alpha \to \infty$, so, if the λ_ℓ's are placed in order of descending real parts the $\lambda_\ell = \alpha\zeta_\ell = c_\ell$ will form the first ℓ_0 distinct eigenvalues. Therefore

$$\lim_{\alpha \to \infty} u_i(t) = \alpha \sum_{\ell=1}^{\ell_0} \sum_{k=0}^{m_\ell - 1} b_{i\ell k} t^k e^{c_\ell t} \qquad (2.2.13)$$

where c_ℓ has finite upper and lower bounds as $\alpha \to \infty$. R is positive definite so, unless $b_{i\ell k} = 0$ for all i,k and for all $\ell \le \ell_0$ $J_\omega \to \infty$ as $\alpha \to \infty$.

The case now left to consider is where $b_{i\ell k}=0$ for all i,k and and for all $t \le t_c$, then

$$J_\alpha = \int_{t_c}^{T} \alpha^2 \sum_{i=1}^{n} r_{ii} \sum_{\ell=\ell_{i1}}^{n_e} \sum_{k=0}^{m_{i\ell}-1} \sum_{\ell'=\ell_{i1}}^{n_e} \sum_{k'=0}^{m_{i\ell'}-1} b_{i\ell k}\, b_{i\ell' k'}\, t^{(k+k')}\, e^{(\lambda_\ell+\lambda_{\ell'})t}\, dt \quad . \tag{2.2.13}$$

It can be seen that we have to consider integrals of the form

$$\int_{t_c}^{T} t^N e^{\gamma t} dt$$

and it is straightforward, using integration by parts, to derive a recursion formula and thence to obtain

$$\int_{t_c}^{T} t^N e^{\gamma t} dt = \frac{e^{\gamma T}}{\gamma}\left[T^N - \frac{N!T^{N-1}}{\gamma(N-1)!} + \frac{N!T^{N-2}}{\gamma^2(N-2)!} \cdots + (-1)^{N-1}\frac{N!T}{\gamma^{N-1}}\right]$$

$$- (-1)^{N-1}\frac{N.!(e^{\gamma T}-1)}{\gamma^{N+1}}$$

where $Re\{\gamma\}<0$ and $0! = 1$. If we let $\gamma \to \alpha(\gamma_c+\epsilon)$ as $\alpha\to\infty$ and ϵ is of the order $1/\alpha$ then it can be seen that

$$\lim_{\alpha\to\infty} \int_{t_c}^{T} \alpha^2 t^N e^{\gamma t} dt = \left\{ \begin{array}{ll} \infty & \text{if } N=0 \\ (-1)^{N-1}\frac{N!}{\gamma_c^{N+1}} & \text{if } N=1 \\ 0 & \text{if } N>1 \end{array} \right\} . \tag{2.2.14}$$

Hence from (2.2.13) and (2.2.14), using the fact that R is positive definite $\lim_{\alpha\to\infty} J_\alpha = \infty$ unless $\sum_{\ell=\ell_{i1}}^{n_e} b_{i\ell 0} = 0$ for all i. However, putting t=0 in (2.2.12) yields

$$u_i(0) = \alpha \sum_{\ell=1}^{n_e} b_{i\ell 0} = 0 \quad \text{for all i}$$

and we have already assumed that $b_{i\ell 0}=0$ for $t \le t_c$, so unless $u_i(0)=0$ for all i, $\lim_{\alpha\to\infty} J_\alpha = \infty$. Therefore the only way that $\lim_{\alpha\to\infty} J_\alpha$ can be finite is for u(0) to be zero and for $u_i(t)$ to contain no terms of the form $b_{i\ell k}\, t^k\, e^{\zeta t}$. Now, u(0) = 0 implies that

$$KCx_c = 0 \text{ , and thence}$$

$$BKCx_c = 0 \text{ .}$$

Hence x_c only contains components in the direction of eigenvectors associated with the zero eigenvalues of BKC. Since the eigenvectors

of $A + \alpha BKC$ are the same as those of $\frac{1}{\alpha}A + BKC$ they can be made arbitrarily close to those of BKC by choosing α sufficiently large. Hence x_0 must have finite components in the direction of the first l_0 eigenvectors of $A + \alpha BKC$. However, as a result of (2.2.13) this is a condition for $J \to \infty$ as then $b_{ik} \neq 0$ for all i,k and for all $l \leq l_0$. Therefore in all the possible cases $\lim_{\alpha \to \infty} J_\alpha = \infty$, so there does exist a realisable optimal control as long as there is some $K \in \mathcal{K}$ that yields a finite value of the cost.

The only other consideration is whether there is a K which leads to a finite cost. One can immediately say that if T is finite J must also be finite since the expression (2.2.11) implies that x(t) is bounded for all finite t. However, if $T \to \infty$ it is quite possible for $J \to \infty$, in general this will be the case if the system is unstable. Systems can be devised in which $J \to \infty$ for all $K \in \mathcal{K}$. An example of such a system is one which is unstable for all $K \in \mathcal{K}$ and in which the initial state vector is never orthogonal to all eigenvectors associated with eigenvalues of F with positive real parts. Hence in these circumstances there is no realisable constrained optimal control. It can be seen that this condition is parallel to the assumption made in chapter 1, section 3, that the system is optimisable relative to Q, in other words that it is possible to construct controls that give a finite cost. In conclusion it may be said that there exists a realisable constrained optimal control for linear time invariant finite dimensional systems if either T is finite or, if T is infinite, there is a $K \in \mathcal{K}$ that yields a finite cost.

Section 3. The calculation of the constrained optimal control.

Having ascertained that a realisable optimal control exists the next problem is to consider how the matrix of feedback gains may be calculated. This has been done by Jameson(1967) who derives the necessary conditions for $\partial J/\partial K_{ij}$ to be zero. His main results using the notation of (2.2.1)-(2.2.3) are as follows.

$$\delta J = \int_0^T tr[C'\delta K'(B'P + RKC)X]dt \qquad (2.3.1)$$

where δJ is the variation in the cost due to a small change δK in the feedback matrix where 2nd order small quantities have been ignored. $X(t)=x(t)x'(t)$ and $P(t)$ is given by

$$\dot{P}(t) + P(t)(A(t)+B(t)KC(t)) + (A(t)+B(t)KC(t))'P(t) + Q(t) + C'(t)K'R(t)KC(t) = 0 \qquad (2.3.2)$$

$$P(T)=G.$$

The operator $tr[.]$ is the trace of a square matrix, that is the sum of the diagonal terms. As has been mentioned earlier, it is reasonable to assume that K is time invariant, then (2.3.1) implies, if δJ is to be zero, that

$$\int_0^T tr[\delta K'(B'P + RKC)XC']dt = 0$$

where the fact that $tr[LM] = tr[ML]$ for all square matrices L,M has been used. If this is to be true for all δK then

$$\int_0^T (B'P + RKC)XC'dt = 0.$$

K can only be written explicitly if R is time invariant, in which case

$$K = -R^{-1}\int_0^T B'PXC' \, dt \left[\int_0^T CXC' \, dt\right]^{-1}. \qquad (2.3.3)$$

If the system itself is time invariant and $T\to\infty$, P becomes time invariant, Athans & Falb(1966), and

$$K = -R^{-1}B'P(CWC')^{-1} \qquad (2.3.4)$$

where

$$W = \int_0^\infty X(t)dt .$$

W is given by

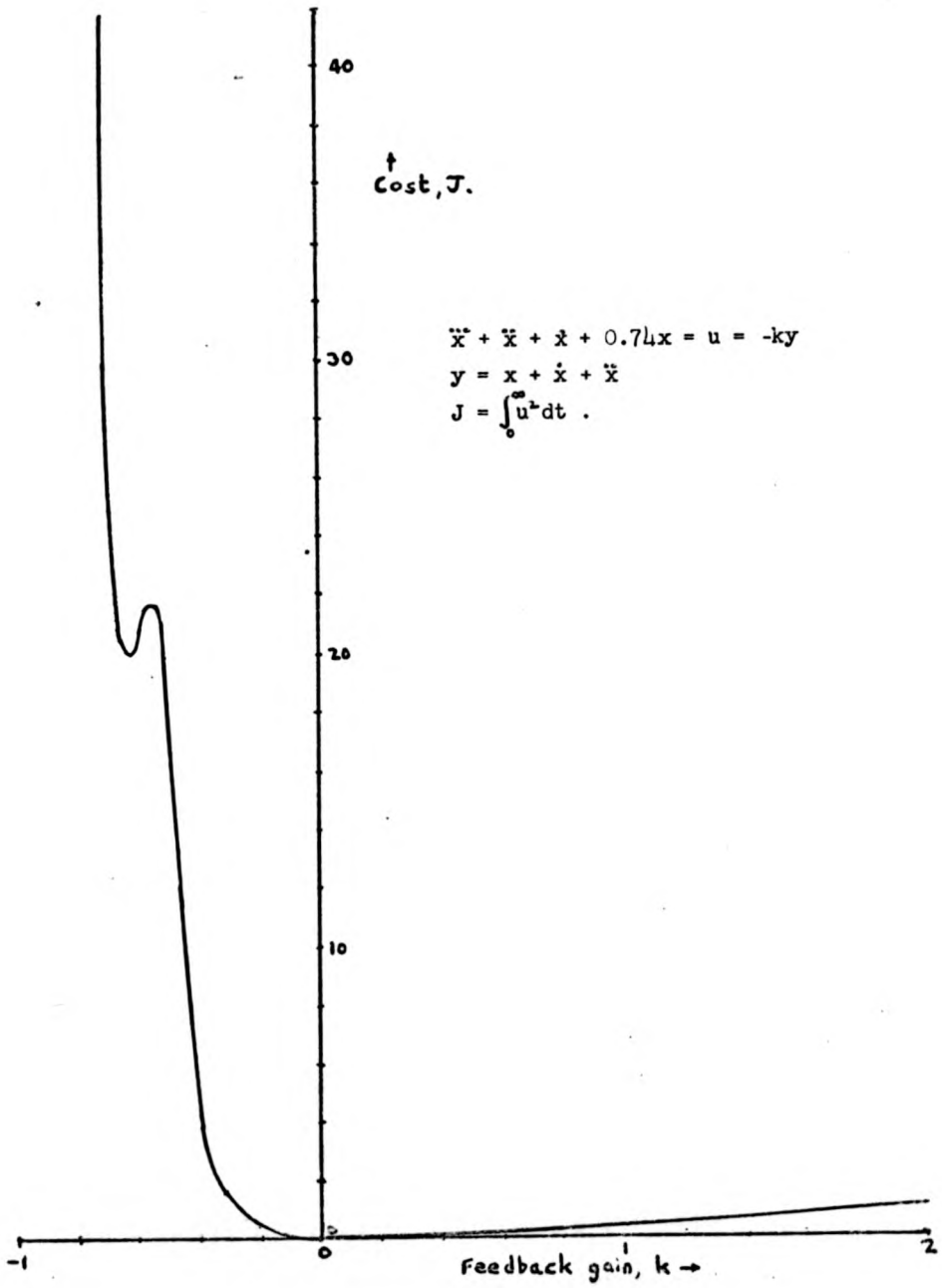
$$(A+BKC)W + W(A+BKC)' + x_o x_o' = 0 \quad . \qquad (2.3.5)$$

On looking at (2.3.2) and (2.3.3) it can be seen that an iterative procedure is possible. If some K is chosen, P(t) can be calculated from (2.3.2) while x(t), and thence X(t), can be derived from (2.2.1)-(2.2.3). These two results, when substituted into (2.3.3) give a new value for K. If the system is time invariant and $T \to \infty$ this method becomes considerably more simple:(2.3.4) and (2.3.5) can then be used , (2.3.2) becomes the simple Liapunov matrix equation

$$P(A+BKC) + (A+BKC)'P + Q + C'K'RKC = 0 \quad , \qquad (2.3.6)$$

hence no differential equations have to be solved in order to carry out the iteration. Unfortunately there seems to be no proof that this process converges, though in practice it often does, nor can one be certain that it will converge to the correct root. Jameson's equations give the points where the gradient of J with respect to the elements of K is zero, so they will also be satisfied at a local maximum as well as at any minimum. Fig.2.1 shows the value of the cost function of a third order system plotted against the single feedback gain, k. It can be seen that this relatively simple system gives rise to quite a complex curve with two minima and one maximum, therefore these equations will be satisfied at three points. Jameson suggests that it may be better to use his expressions for the gradients in some search procedure rather than trying to solve the equations directly. This will obviate the possibility of finding a maximum, but it does not guarantee that the lowest local minimum will be found.

Fig. 2.1


40
30
20
10
0
Cost, J.
$\dddot{x} + \ddot{x} + \dot{x} + 0.74x = u = -ky$
$y = x + \dot{x} + \ddot{x}$
$J = \int_0^\infty u^2 dt$.
-1
0
1
2
Feedback gain, k →

However, there is a way of modifying the direct iteration procedure to obtain a new method which is guaranteed to generate a sequence of controllers each of which reduces the cost compared with the previous one; this we shall call the fractional step algorithm. Each step technically has to be infinitesimal, so it is not as powerful a method as that described in Chapter 1 for calculating the totally observed optimal control.

The direct iteration method involves calculating P from K and then using (2.3.3) to calculate the next value of K. Now consider the consequences of just moving K part of the way towards the value indicated by (2.3.3) . If $K = K_i$ at the i^{th} iteration set

$$K_{i+1} = K_i + \left\{-R^{-1}\int_0^T B'PXC'dt.\left[\int_0^T CXC'dt\right]^{-1} - K_i\right\} \quad (2.3.7)$$

where $\alpha \ll 1$. The change in K, δK, is then given by

$$\delta K = -\alpha\left\{R^{-1}\int_0^T B'PXC'dt.\left[\int_0^T CXC'dt\right]^{-1} + K_i\right\}$$

which for clarity can be written

$$\delta K = -\alpha\left\{R^{-1}YZ + K_i\right\} . \quad (2.3.8)$$

Since α is small we may use (2.3.8) to calculate the first order change in cost

$$\delta J = -\alpha tr\left[(R^{-1}YZ + K_i)'(Y + RK_iZ^{-1})\right]$$
$$= -\alpha tr\left[Z(R^{-1}Y + K_iZ^{-1})'R(R^{-1}Y + K_iZ^{-1})\right]. \quad (2.3.9)$$

Z is a positive definite, symmetric matrix so we may write

$$Z = Z^{\frac{1}{2}} Z^{\frac{1}{2}}$$

and thus

$$\delta J = -\alpha tr\left[Z^{\frac{1}{2}}(R^{-1}Y + K_iZ^{-1})'R(R^{-1}Y + K_iZ^{-1})Z^{\frac{1}{2}}\right]. \quad (2.3.10)$$

R is positive definite so, provided $\alpha > 0$, δJ must be negative. Therefore the iteration scheme given by (2.3.7) generates a new controller that leads to a first order reduction in the cost, as long as α is small compared with unity.

In the practical application of this method it is obviously going to be necessary to choose some finite value of α. The smaller α is the more certain is the iteration to converge but the longer it will take, so some compromise has to be found. This problem is considered for a practical example in Chapter 5; a further improvement is introduced there in which α is initially chosen quite large but if any element of K changes by more than a certain fraction α is correspondingly reduced; this appears to be successful. The fractional step algorithm is guaranteed to converge to a minimum but it does not overcome the problem of multiple minima as shown in Fig.2.1. The only way of overcoming such difficulties seems to be to initiate the search from different points and check that the final answers are all the same.

In conclusion we may say that the methods discussed here of equating the derivative of the cost to zero can be very useful in computing the constrained optimal control, however, the results must be checked to see whether the correct minimum has been found. It may turn out to be better to use a search routine to minimise J directly, in which case Jameson's equations can be used to give an explicit expression for the gradient of J. The fractional step algorithm, though, appears the most promising as it is certain to give convergence to a local minimum.

Section 4. Bounds on the cost function.

The calculation of both the constrained and unconstrained optimal control can be very time consuming, therefore it would be useful if bounds on the cos can be found with less effort. The optimal control with complete knowledge of the state gives the global minimum of the cost so we should like to compare any other control with this. A method will be presented which avoids calculation of the optimal control and moreover can be applied to infinite dimensional systems satisfying the conditions of Chapter 1.

The dynamics of the system are described, as in Chapter 1, by

$$\dot{z}(t) = Az(t) + B\ u(t)\ ,\quad z(0) = z_0 \qquad (2.4.1)$$

where the state $z(t)$ and the control $u(t)$ are elements of Hilbert spaces H and U respectively. A must be a closed linear operator defined on a dense domain $D(A) \subseteq H$ that also is the infinitesimal generator of a strongly continuous semigroup, T_t ; B is a bounded linear operator. A mild solution to (2.4.1) may then be defined as in (1.1.4), that is

$$z(t) = T_t z_0 + \int_0^t T_{t-\sigma} B\ u(\sigma)d\sigma\ . \qquad (2.4.2)$$

The cost functional of (1.1.8) is again associated with the control problem

$$J(u) = \langle z(T), Gz(T)\rangle_H + \int_0^T \left[\langle z(t),\ Q\ z(t)\rangle_H + \langle u(t),\ R\ u(t)\rangle_U\right] dt. \qquad (2.4.3)$$

In this expression Q is a bounded self-adjoint positive definite operator on H, G is a bounded self-adjoint positive semidefinite operator on H and R, R^{-1} are bounded self adjoint operators on U. Hence there exist $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \beta_4$ all positive such that, for all $t\in[0,T]$,

$$\alpha_1\|z\|_H^2 \geqslant \langle z,\ Q\ \ z\rangle_H \geqslant \alpha_2\|z\|_H^2$$

$$\alpha_3\|z\|_H^2 \geqslant \langle z, Gz\rangle_H \geqslant 0$$

$$\beta_1\|u\|_U^2 \geqslant \langle u,\ R\ \ u\rangle_U \geqslant \beta_2\|u\|_U^2$$

$$\beta_3\|u\|_U^2 \geqslant \langle u,\ R^{-1}\ u\rangle_U \geqslant \beta_4\|u\|_U^2.$$

It may be noted that the only extra assumption made over those in Chapter 1 is that Q is confined to be positive definite rather than positive semidefinite. However, it will be shown that this condition may be relaxed in certain circumstances.

We shall now proceed to determine the bounds on J(u). The feedback control under consideration will be taken as

$$u_0(t) = -\ R^{-1}\ B^*\ \ P_0(t)z(t) = K_0(t)z(t) \qquad (2.4.5)$$

and this will be compared with any other control, u(t), which may indeed be the optimal control $u^*(t)$. The mild solution of (2.4.1)

and (2.4.5) are given by (2,4,2) to be

$$z(t) = T_t z_o - \int_o^t T_{t-\sigma} B R^{-1} B^* P_o(\sigma) d\sigma \qquad (2.4.6)$$

or, equivalently from (1.1.6), using the evolution operator $S(t,\sigma)$ with $E(t) = -B R^{-1} B^* P_o(t)$

$$z(t) = S(t,0) z_o .$$

The objective is to find a bound μ such that

$$\frac{J(u)}{J(u_o)} \geqslant \mu \qquad (2.4.7)$$

or, if $u=u^*$, the optimal control,

$$1 \geqslant \frac{J(u^*)}{J(u_o)} \geqslant \mu . \qquad (2.4.8)$$

The lower bound μ depends on u_o so it will give an indication of the effectiveness of u_o. The nearer μ is to unity the better the control u_o as then the margin for improvement is smaller. Moreover, μ may be used as a cost criterion, that is u_o can be chosen to maximise μ instead of minimising $J(u_o)$. This corresponds to one of the methods for dealing with incomplete feedback when the initial state is unknown given in chapter 3 and by Levine, Johnson and Athans(1971). The method involves finding the constrained control that minimises the maximum over all initial states of $J(u)/J(u^*)$ and the direct calculation of this necessitates knowing u^*. Since $J(u)/J(u^*) \leq 1/\mu$ for all initial states an estimate for such a criterion will be obtained by maximising μ and has the advantage that u^* need not be calculated.

In order to calculate the lower bound let

$$A_o(t) = A - B R^{-1} B^* P_o(t)$$

and $S(t,\sigma)$ be the evolution operator generated by A_o. Set

$$P(t) = S^*(T,t) G S(T,t) + \int_t^T S^*(\sigma,t) \left[Q + P_o(\sigma) B R^{-1} B^* P_o(\sigma) \right] S(\sigma,t) d \qquad (2.4.9)$$

and consider the control system

$$\dot{z}(t) = A_o z(t) + B\,\bar{u}(t) \quad . \tag{2.4.10}$$

Hence the actual control for this system is

$$u(t) = -R^{-1} B^* P_o(t)\, z(t) + \bar{u}(t) \tag{2.4.11}$$

and the mild solution of (2.4.10) is

$$z(t) = S(t,0)z_o + \int_o^t S(t,\sigma)\, B\, \bar{u}(\sigma)\, d\sigma \quad . \tag{2.4.12}$$

We shall need some of the theorems developed in Chapter 1 and since theorem I plays such a central role in the analysis it is worthwhile to give some idea of the thinking that leads to the result. This can easily be illustrated by considering the finite dimensional case when the operator A is bounded. The method of proof for the infinite dimensional case is necessarily more complex as it involves using the mild integral forms (2.4.9), (2.4.12) whereas in the finite dimensional system the integrals need no longer be used. The strong forms of the equations for z(t) and P(t) are

$$\dot{z}(t) = A_o(t)z(t) + B\bar{u}(t) \tag{2.4.13}$$

$$\dot{P}(t) + A_o'(t)P(t) + P(t)A_o(t) + Q + P_o(t)BR^{-1}B'P_o(t) = 0 \tag{2.4.14}$$

$$P(T) = G \quad .$$

Let $V(t) = z'(t)P(t)z(t)$ so

$$\dot{V}(t) = \dot{z}'(t)P(t)z(t) + z'(t)P(t)\dot{z}(t) + z'(t)\dot{P}(t)z(t)$$

$$= z'(t)\left[A_o'(t)P(t) + P(t)A_o(t) + \dot{P}(t)\right]z(t) \tag{2.4.15}$$

$$+ \bar{u}'(t)B'P(t)z(t) + z'(t)P(t)B\bar{u}(t) \quad .$$

Integrating (2.4.15) from t to T and using (2.4.14)

$$V(T) - V(t) = \int_t^T \left\{-z'(\sigma)\left[Q + P_o(\sigma)BR^{-1}B'P_o(\sigma)\right]z(\sigma) + 2\bar{u}'(\sigma)B'P(\sigma)z(\sigma)\right\}d\sigma \quad .$$

Now, $V(T) = z'(T)Gz(T)$, hence

$$z'(t)P(t)z(t) = V(t) = z'(T)Gz(T) + \int_t^T \left\{z'(\sigma)\left[Q + P_o(\sigma)BR^{-1}B'P_o(\sigma)\right]z(\sigma) - 2\bar{u}'(\sigma)B'P(\sigma)z(\sigma)\right\}d\sigma \quad . \tag{2.4.16}$$

This last equation is equivalent to the result of theorem I in chapter 1.

This theorem will now be used to obtain the lower bound .

Putting t=0 into theorem I and using (2.4.9) gives

$$J(u) - J(u_0) = \int_0^T \left\{ \langle \bar{u}(\sigma), R\ \bar{u}(\sigma) \rangle_U + 2\langle z(\sigma), [P(\sigma)-P_c(\sigma)]\ B\ \bar{u}(\sigma) \rangle_H \right\} d\sigma . \tag{2.4.17}$$

It will be shown that if a $\gamma > 0$ can be found such that

$$\gamma \langle z(t), Q\ z(t) \rangle_H + \frac{\gamma}{1+\gamma} \langle z(t), P(t) B\ R^{-1}\ B^*\ P(t)\ z(t) \rangle_H$$

$$\geq \langle [P(t)-P_c(t)]\ z(t),\ B\ R^{-1}\ B^*\ [P(t)-P_c(t)]\ z(t) \rangle_H \tag{2.4.18}$$

for all $t \in [0,T]$, then

$$J(u)-J(u_0) \geq -\gamma \int_0^T \left[\langle u(t), R\ u(t) \rangle_U + \langle z(t), Q\ z(t) \rangle_H \right] dt$$

$$= -\gamma J(u) . \tag{2.4.19}$$

$R(t)$ is a positive definite operator and we have assumed that $\gamma > 0$ so the inequality (2.4.18) will still hold if we add

$$(1+\gamma)\langle \bar{u}(t) + R^{-1}\ B^*\ \left[\frac{P(t)}{1+\gamma} - P_c(t)\right] z(t),\ R \left\{ \bar{u}(t) + R^{-1}\ B^*\ \left[\frac{P(t)}{1+\gamma} - P_c(t)\right] z(t) \right\rangle_H$$

to the left hand side, this gives

$$(1+\gamma)\left\langle \bar{u} + R^{-1}B^*\left[\frac{P}{1+\gamma} - P_c\right]z, R\left\{\bar{u} + R^{-1}B^*\left[\frac{P}{1+\gamma} - P_c\right]z\right\}\right\rangle_U + \gamma\langle z, Qz\rangle_H + \frac{\gamma}{1+\gamma}\langle z, PBR^{-1}B^*Pz\rangle_H$$

$$\geq \langle (P-P_c)z, BR^{-1}B^*(P-P_c)z \rangle_H \tag{2.4.20}$$

where the dependence of the variables on t has not been shown in order to clarify the manipulations of the inequalities. Hence

$$(1+\gamma)\langle \bar{u}, R\bar{u}\rangle_U + 2\left\langle \bar{u}, B^*\left[P-(1+\gamma)P_c\right] z\right\rangle_U + \frac{1}{1+\gamma}\left\langle \left[P-(1+\gamma)P_c\right] z, BR^{-1}B^*\left[P-(1+\gamma)P_c\right]z \right\rangle_H$$

$$+\gamma\langle z, Qz\rangle_H + \frac{\gamma}{1+\gamma}\langle z, PBR^{-1}B^*Pz\rangle_H \geq \left\langle (P-P_c)z, BR^{-1}B^*(P-P_c)z \right\rangle_H$$

which simplifies to

$$(1+\gamma)\langle \bar{u}, R\bar{u}\rangle_U + 2\langle \bar{u}, B^*Pz\rangle_U - 2\langle \bar{u}, B^*P_c z\rangle_U \geq -\gamma\langle z, Qz\rangle_H + 2\gamma\langle \bar{u}, B^*P_c z\rangle_U - \gamma\langle z, PBR^{-1}B^*Pz\rangle_H .$$

Therefore, on rearranging and integrating from 0 to T

$$\int_0^T \left[\langle \bar{u}, R\bar{u} \rangle_U + 2\langle \bar{u}, B^*(P-P_0)z \rangle_U \right] dt \geq$$

$$-\gamma \int_0^T \left[\langle z, Qz \rangle_H + \langle -R^{-1}B^*P_0 z+\bar{u}, R(-R^{-1}B^*P_0 z+\bar{u}) \rangle_U \right] dt$$

$$= -\gamma \int_0^T \left[\langle z, Qz \rangle_H + \langle u, Ru \rangle_U \right] dt = -\gamma J(u)$$

from (2.4.5). Hence from (2.4.17)

$$J(u) - J(u_0) \geq -\gamma J(u), \quad \frac{J(u)}{J(u_0)} \geq \frac{1}{1+\gamma} = \mu \ . \qquad (2.4.21)$$

We must now consider the conditions that must be met for (2.4.18) to be satisfied by some $\gamma>0$. Q, P_0, B, R and G are uniformly bounded operators and from (2.4.9) it is straightforward to show that $P(t)$ is also uniformly bounded, thus the problem becomes: find a $\gamma>0$ such that

$$a + b.\frac{\gamma}{1+\gamma} \geq c \qquad (2.4.22)$$

where $a, b, c > 0$. If $\gamma>0$

$$0 < \frac{\gamma}{1+\gamma} < 1 \ ,$$

so, provided $a>0$, it is always possible to find a γ such that (2.4.22) is satisfied. Since it has been assumed that Q is positive definite a will always be positive, hence a suitable γ can always be found. If, however, $a=0$, that is Q is only positive semi-definite, then there is a $\gamma>0$ satisfying (2.4.22) only if $b>c$. This condition, when written in full, becomes

$$\langle z, PBR^{-1}B^*Pz \rangle_H > \langle z, (P-P_0)BR^{-1}B^*(P-P_0)z \rangle_H \ .$$

This will be true if the operator

$$P(t)B(t)R^{-1}(t)B^*(t)P(t) - \left[P(t)-P_0(t)\right] B(t)R^{-1}(t)B^*(t)\left[P(t)-P_0(t)\right] \qquad (2.4.23)$$

is positive definite for all $t\in[0,T]$. If $P(t)$ is close to $P_0(t)$ in some sense then γ will be small and the cost $J(u_0)$ will be near the minimimal cost $J(u^*)$. Also in this case it is more likely that

the operator in (2.4.23) will be positive definite, in which case the condition that Q must be positive definite may be relaxed. Since, when Q is positive semidefinite, a is only zero if $z = \bar{z}$, an eigenfunction corresponding to a zero eigenvalue of Q, then, as $P_o(t) \to P(t)$, (2.4.22) will be satisfied by (2.4.23) being only positive semidefinite provided that $\langle \bar{z}, PBR^{-1}B^*P\bar{z} \rangle_4 \neq 0$. Indeed, if $P(t)=P_o(t)$ for all $t\epsilon[0,T]$ then $P_o(t)$ must satisfy a mild version of the Riccati equation and the control generated by $P_o(t)$ will be the optimal, $u^*(t)$. Alternatively, if the control is constrained in some manner, the free parameters of $P_o(t)$ can be chosen in such a way as to minimise γ. These results may be extended to the infinite time interval case by replacing the functional (2.4.3) by

$$J(u) = \int_0^\infty \left[\langle z(t), Qz(t) \rangle_4 + \langle u(t), Ru(t) \rangle_3 \right] dt.$$

We shall now show how these bounds on the value of the cost can be applied to some examples consisting of both finite and infinite dimensional systems.

Example 1.

Consider the damped oscillator

$$\ddot{x} + \dot{x} + x = u$$

where the control u is linearly dependent on the displacement only, that is

$$u = -kx$$

and the velocity, $\dot{x}$, cannot be measured.

The cost function is taken to be

$$J(u) = \int_0^\infty (x^2 + \dot{x}^2 + u^2) dt.$$

Thus

$$A = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 1$$

$$P_o = \begin{bmatrix} a & k \\ k & 0 \end{bmatrix}$$

where a is arbitrary and

$$P = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix}$$

where

$$p_2 = \frac{1+k^2}{2(1+k)} \quad , \quad p_3 = \tfrac{1}{2} + p_2 \quad , \quad p_1 = p_2 + (1+k)p_3 \quad .$$

These elements of P are all time invariant since the period of integration in the cost function is infinite and the coefficients in the differential equation governing the system are constant over time. (2.4.18) will be satisfied if

$$\gamma(x^2 + \dot{x}^2) + \frac{\gamma}{1+\gamma}(p_1^2 x^2 + 2p_2 p_3 x\dot{x} + p_3^2 \dot{x}^2) \geqslant (p_2 - k)^2 x^2 + 2p_3(p_2 - k)x\dot{x} + p_3^2 \dot{x}^2 \quad .$$

As P is not a simple function of k we cannot easily find the k that minimises γ, however, it is straightforward to compare any two controls. First consider the uncontrolled system, k=0. Set P_o=0 then (2.4.18) gives that $\gamma(1+\gamma)$ is the maximum eigenvalue of $PBR^{-1}B^*PQ^{-1}$ if Q is positive definite. Therefore in this example we obtain

$$1 \geqslant \frac{J(u^*)}{J(0)} \geqslant 0.58.$$

If k$\neq$0 we have first to calculate P and then examine the inequality (2.4.22) to find γ. Putting k=$\frac{1}{2}$ and using this procedure the following bounds result

$$1 \geqslant \frac{J(u^*)}{J(u^-)} \geqslant 0.67 \quad .$$

Hence it may be seen that according to the criterion used here the feedback control k=$\frac{1}{2}$ is better than that where k=0. When compared with the optimal cost $J(u^*)$ 33% improvement is possible for k=$\frac{1}{2}$ whereas 42% improvement can take place over the uncontrolled system. This procedure can be carried out for any value of k>-1, the feedback

which ensures finite cost, until the control is found that minmises the possible improvement.

Example 2.

Consider the distributed parameter system representing diffusion

$$\frac{\partial z(t,x)}{\partial t} = \frac{\partial^2 z(t,x)}{\partial x^2} + u(t,x) \quad , \quad x \in [0,1] \ ,$$

subject to the boundary conditions

$$z(t,0) = z(t,1) = 0 \ ,$$

and the initial condition

$$z(0,x) = z_c(x) \ .$$

The cost functional is taken to be

$$J(u) = \int_0^\infty \int_0^1 \left[z^2(t,x) + \lambda u^2(t,x) \right] dx \, dt$$

where $\lambda > 0$. The uncontrolled system is asymptotically stable with respect to the norm

$$\left[\int_0^1 z^2(t,x) dx \right]^{1/2}$$

so that the cost is finite and can be compared with the cost of any other control. The operator P which determines the cost when u=0 is obtained from (2.4.9) with $P_c(t)=0$ and is assumed to be given by the integral operator

$$Pz = \int_0^1 p(x,y) z(t,y) dy \ ,$$

and P is time invariant since we are considering the infinite interval. Then p(x,y) must satisfy

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \delta(x - y) = 0$$

$$p(x,0) = p(x,1) = 0 \quad \text{for all } x \in [0,1]$$

$$p(0,y) = p(1,y) = 0 \quad \text{for all } y \in [0,1]$$

where $\delta(x-y)$ is the Dirac delta function. Hence

$$p(x,y) = \sum_{m=1}^{\infty} \frac{1}{m^2 \pi^2} \sin m\pi x . \sin m\pi y \ .$$

Since B and Q are the identity operators (2.4.18) is satisfied by finding the smallest value of γ for which

$$\int_0^1 \left[\int_0^1 p(x,y)z(t,x)dx\right]^2 dy \leq \gamma(\gamma+1)\int_0^1 z^2(t,x)dx \ .$$

By the Schwarz inequality

$$\left[\int_0^1 p(x,y)z(t,x)dx\right]^2 \leq \int_0^1 p^2(x,y)dx \int_0^1 z^2(t,x)dx \ .$$

Hence

$$\lambda\gamma(\gamma+1) \geq \int_0^1\int_0^1 p^2(x,y)dxdy = \frac{1}{4\pi^4}\sum_{m=1}^{\infty}\frac{1}{m^4} = \frac{1}{360} \ .$$

Thus

$$1 \geq \frac{J(u^*)}{J(0)} \geq 180\lambda\left[\sqrt{1 + 1/90\lambda} - 1\right] \ .$$

Alternatively, since P is a compact operator, its spectrum will consist of a countable number of eigenvalues with the only point of accumulation being at zero. In particular the maximum eigenvalue is $1/2\pi^2$, hence

$$\left\|\int_0^1 p(x,y)z(t,x)dx\right\| \leq \frac{1}{2\pi^2}\|z\| \ .$$

Therefore

$$\lambda\gamma(\gamma+1) = 1/4\pi^4 \ , \text{ and}$$

$$1 \geq \frac{J(u^*)}{J(0)} \geq 2\pi^4\lambda\left[\sqrt{1 + 1/\pi^4\lambda} - 1\right] \ .$$

For any given λ this gives a slightly higher, and thus better, estimate for γ. Such a result would be expected since the eigenvalue derived upper limit must give the true supremum of $\|Pz\|$. Looking at the results it can be seen that if $\lambda > 1$ the best controller improves on no control by at most 1%. On the other hand, for small λ, when the penalty for using too much control action is reduced, then, as one would expect, it may be desirable to construct controllers. For example, if $\lambda = 0.001$, $\mu \doteq 0.45$, indicating that the cost could possibly be more than halved by introducing a controller.

If the control is confined to the boundary so that

$$z(t,0) = u(t),\ z(t,1) = 0\ ,$$

and where the performance index is given by

$$J(u) = \int_0^\infty \left[\int_0^1 z^2(t,x)dx + \lambda u^2(t) \right] dt\ ,$$

the analysis is not applicable since the operator B will be unbounded and the performance index J(u) may not make sense. However one may proceed formally and and obtain results in the same manner as before. Consequently the following inequality is derived

$$1 \geq \frac{J(u^*)}{J(0)} \geq 6\lambda\left[\sqrt{1 + 1/3\lambda} - 1\right] .$$

This has the same form as the previous example with distributed control but the values of γ are lower for the same value of λ. For instance, if $\lambda=1$, $\gamma=0.9$, so up to 10% improvement may be possible. We may again deduce that the desirability of constructing controllers increases as λ decreases.

Example 3.

In this last example control action is confined to the boundary and the uncontrolled state is not asymptotically stable. The system is

$$\frac{\partial z(t,x)}{\partial t} = \frac{\partial^2 z(t,x)}{\partial x^2}$$

$$\frac{\partial z(t,0)}{\partial x} = u(t)\ , \quad \frac{\partial z(t,1)}{\partial x} = 0$$

$$z(0,x) = z_0(x)$$

and

$$J(u) = \int_0^\infty \left[\int_0^1 z^2(t,x)dx + \lambda u^2(t) \right] dt\ .$$

The simplest type of feedback control which stabilises such a system and gives finite values for the performance index is one based on a single sensor at the point $x=a$ so that $u(t)=Gz(t,a)$ where G is

is constant; see Parker(1970) for the proof of the stability of this system. However, not only is this controller a mathematically idealised description of the practical situation, no sensor is of infinitesimal size measuring something at a single point, but also the earlier analysis is not strictly applicable as B is again unbounded. As well as this Gz(t,a) cannot be bounded by the norm

$$\|z\| = \left[\int_0^1 z^2(t,x)dx\right]^{1/2} .$$

In this example we shall circumvent these problems, apart from the unboundedness of B, by taking the control to be of the form

$$u_c(t) = G\int_0^1 g(x)z(t,x)dx$$

where g(x) is a continuous function, non-negative on some finite interval containing the point x=a, zero outside that interval and such that

$$\int_0^1 g(x)dx = 1 .$$

This kind of control allows for the sensor to take up some average value of the variable z over a finite interval, and it is also immediately clear that $\int_0^1 g(x)z(t,x)dx$ is bounded with respect to $\|z\|$. Moreover, this way of describing the sensing not only makes the problem tractable mathematically, but also is a much truer description of any physical method of measurement.

As before, when dealing with an unbounded B, we may proceed formally in order to evaluate the bounds on the cost. Assume the operator P that gives the cost of the control $u_c(t)$ to be of the form

$$Pz = \int_0^1 p(x,y)z(t,y)dy .$$

Then p(x,y) must satisfy

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + G\, g(x)g(y) - Gg(y)p(x,0) - Gg(x)p(0,y) = 0 ,$$

$$\frac{\partial p}{\partial x}(0,y) = \frac{\partial p}{\partial x}(1,y) = 0 \text{ for all } y,$$

$$\frac{\partial p}{\partial y}(x,0) = \frac{\partial p}{\partial x}(x,1) = 0 \text{ for all } x.$$

In order to obtain the lower bound for the ratio of the cost of any other control to the cost of $u_0(t)$ it is necessary to find the smallest value of γ for which

$$\gamma \int_0^1 z^2(t,x)dx + \frac{\gamma}{\lambda(\gamma+1)}\left[\int_0^1 p(0,y)z(t,y)dy\right]^2 \geqslant \frac{1}{\lambda}\left[\int_0^1 \{p(0,y) - \lambda G g(y)\}\ z(t,y)dy\right]^2 .$$

An upper bound for this value is obtained via the Schwarz inequality and is given by

$$\gamma = \frac{1}{\lambda}\int_0^1 [p(0,y) - \lambda G g(y)]^2 dy .$$

For λ=0.01 and g(x) of the form shown in Fig.2.2 the best values of a,G which minimise γ can be obtained by numerical search routines and are

$$a = 0.12, \quad G = 6.0, \quad \gamma = 0.85.$$

Hence

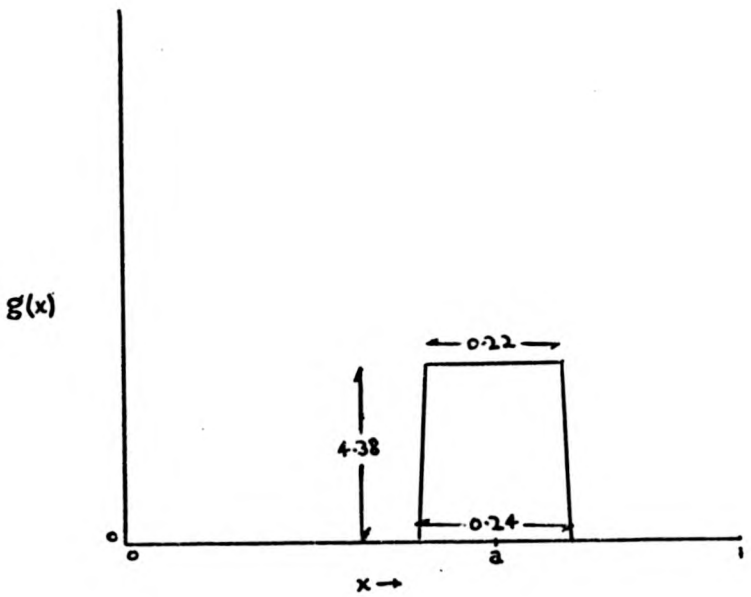
$$1 \geqslant \frac{J(u^*)}{J(u_0)} \geqslant 0.54 .$$

The choice of g(x) is somewhat arbitrary and has no theoretical or practical jutification, though the form chosen does not seem unreasonable. The whole question of the design and positioning of sensors needs further investigation. There are two main aspects.

(i) The determination of the way in which a sensor of finite size averages the variable.

(ii) In the design of a sensor, certain parameters, such as size and shape, may be open to choice. These parameters will affect the averaging process and will therefore change the control and the cost. The design of sensors should therefore play as important a part in the optimisation process as the determination of their positions and associated gains.

Fig. 2.2



It can be seen that this method of obtaining bounds on the cost is very useful for comparing two controls. For instance, in the first part of Example 2 it turns out that the uncontrolled system can be practically as good as the optimally controlled system, so a lot of time and effort can be saved by realising that it is not worth building a controller at all. The results give an idea of how much improvement is possible by using feedback from all states and attaining the theoretical optimal control; there is also a great advantage in being able to find such a bound without actually computing the optimal feedback . However, it should be remembered that this result tells one that the cost cannot be worse than a certain value, whereas in fact the bound may be unduly pessimistic. Therefore the ranking of a set of controls by this method does leave room for doubt. Using the bound as a criterion for controller design is best applied to relatively simple systems where only a few parameters are open to choice. The reason for this is that the minimisation of γ with respect to the feedback gains has to be done by a direct search technique which can become very time consuming when many variables can be changed. Therefore it could be better to use Jameson's equations for the gradients in conjunction with some of the methods presented in chapter 3 to find a control that is "best" in some sense.

CHAPTER 3

THE APPLICATION OF THE CONSTRAINED OPTIMAL CONTROL WITH UNKNOWN INITIAL STATE

Section 1. Introduction.

Chapter 1 dealt with the derivation of the optimal control and it was shown that under certain conditions the optimal control did exist for both finite and infinite dimensional systems. In chapter 2 it was pointed out that this control required complete knowledge of the state of the system at all times which, in practice, would be difficult to implement for all but the simplest systems. Jameson's work was commented on and it turned out that if a constrained optimal control existed, it would depend on the initial state. However, since the purpose of building a feedback control is to make the system return to its equilibrium point when subjected to a wide range of disturbances it is unlikely that the initial state would be known exactly. Thus the problem becomes one of defining what is the best constrained controller when the initial state is unknown.

Several proposals have been made regarding this problem of which some are given by Rekasius(1967), Kleinman & Athans(1968), Jameson(1970) and Levine, Johnson & Athans(1971). We shall present the problem in finite dimensions as the concept of having fewer observable outputs than state variables is more comprehensible than in an infinite dimensional system. Moreover, as has been stated earlier, numerical calculations performed on infinite dimensional systems are almost certain to involve some finite approximation. On the other hand, the criteria relating the cost to some other quadratic form in the initial state are valid for any type of system.

Section 2. Design criteria with unknown initial state.

Consider the linear finite dimensional system described by equations (2.2.1)-(2.2.4). The cost is then given by

$$J = x_o' P(0) x_o \quad , \tag{3.2.1}$$

where P(t) is given by (2.3.2). We may write the expression for J, (3.2.1), in an alternative form, namely

$$J = tr\left[P(0)X_o\right] \tag{3.2.2}$$

where $X_o = x_o x_o'$ and $tr[.]$ designates the trace of a matrix.

If x_o is known, the constrained optimal control problem is to find

$$\inf_K[J] = \inf_K\left[x_o' P(0) x_o\right] \quad , \tag{3.2.3}$$

a problem that can be approached using the methods of Jameson detailed in chapter 2. It is quite possible that although the initial state is not known precisely one does have some prior knowledge of its probability density distribution; in this case the form (3.2.2) can be useful. The simplest approach using the statistics of x_o is to minimise the expected value of J, so, taking expected values in (3.2.2)

$$E\{J\} = E\{tr\left[P(0)X_o\right]\} \quad .$$

The trace operator is linear and P(0) is independent of x_o so

$$E\{J\} = tr\left[P(0)E\{X_o\}\right] \quad . \tag{3.2.4}$$

Hence if $E\{X_o\}$ is known $E\{J\}$ can be minimised with respect to K. Since the expected value operation is linear Jameson's equations are still useful for a time invariant system considered over the infinite interval. If expected values are taken in (2.3.1), one infers that for $E\{J\}$ to be a minimum X(t) must be replaced by $E\{X(t)\}$. This being true since all the other terms on the right hand side are

independent of x_o. It can be seen that as a result of taking expected values in(2.3.4) and (2.3.5) it is only necessary to replace $X_o$by $E\{X\}$ in (2.3.5) in order to minimise $E\{J\}$; W will now be equal to $\int_o^\infty E\{X(t)\}\,dt$. In fact Levine & Athans(1970) have proved this result, using methods very similar to those of Jameson, for the case where $E\{X_o\}=I$, the unit matrix. The finite interval time varying case is more difficult to deal with since one has then to calculate $E\{X(t)\}$ for all $t\epsilon[0,T]$. The optimisation criterion put forward by Levine & Athans is one quite commonly postulated as it gives rise to a relatively simple problem of minimising $tr[P(0)]$. The inherent assumption underlying this is that x_o has a probability density function such that it is uniformly distributed over the surface of the unit sphere.

The other criteria we shall consider are those of the "worst case" type involving the solution of a min-max problem. Typically this consists of finding the initial state that gives the ratio of the cost to some other quadratic form in the initial state its maximum, or worst case, value. Hence the general problem is to find

$$\min_{K} \max_{x_o} (\mu) \qquad (3.2.5)$$

where

$$\mu = \frac{x_o' P(0) x_o}{x_o' S x_o} \qquad (3.2.6)$$

and it is assumed that $x_o \neq 0$. S in an n×n symmetric matrix that must be positive definite or else there exists an $x_o \neq 0$ for which $\mu \to \infty$. The reason for choosing μ of this form is that suitable choices of S lead to criteria which one can readily see to be of practical significance. Also, by choosing the ratio of two quadratic forms μ is independent of the absolute scale of x_o. If S

is made equal to I, the unit matrix, the objective expressed in (3.2.5) and (3.2.6) is to find the K which minimises the maximum possible cost that can occur for any initial state with a given norm $\|x_0\| = \sqrt{x_0^! x_0}$. Alternatively, if $S = P^*$ the cost matrix associated with the optimal control when one has complete knowledge of the state, μ is the the ratio of the actual cost to the optimal cost. Hence one would find the control that minimises the maximum possible fractional increase in cost that can occur as a result of incomplete feedback. An account of these criteria is also given in Levine, Johnson & Athans(1971).

If there are no constraints on x_0 differentiation of μ with respect to x_0 gives

$$\frac{\partial \mu}{\partial x_0} = \frac{2(x_0^! S x_0) P(0) x_0 - 2(x_0^! P(0) x_0) S x_0}{(x_0^! S x_0)^2} .$$

At a stationary value of μ $\partial\mu/\partial x_0 = 0$, so

$$P(0)x_0 - \frac{x_0^! P(0) x_0}{x_0^! S x_0} S x_0 = P(0)x_0 - \mu S x_0 = 0 .$$

S is positive definite so S^{-1} exists and

$$P(0)S^{-1}(S x_0) = \mu(S x_0) .$$

Therefore μ must be an eigenvalue of $P(0)S^{-1}$ and hence it follows that the maximum value of μ is the maximum eigenvalue of $P(0)S^{-1}$.

Jameson(1970) shows how these criteria can be incorporated into his equations by replacing x_0 with the eigenvector of $P(0)S^{-1}$ corresponding to the maximum eigenvalue, his paper also includes some numerical examples. It should be pointed out that choosing $S = P^*$ has the disadvantage that the calculation of P^* is relatively difficult and it may be preferable to use the methods presented in chapter 2 for finding bounds on the ratio μ.

We shall now show how, according to this criterion, better controls

may be constructed by noticing that one does have some knowledge of x_o. It is assumed that y(t) is known for all t so the measurement of y(0) yields some information about x through the equation

$$Cx_o = y(0) \ . \qquad (3.2.7)$$

It is possible to derive (m-1) homogeneous constraints from (3.2.7),

$$C_o x_o = 0 \qquad (3.2.8)$$

where C_o depends on C and y(0). These, together with any one of the original constraints in (3.2.7)

$$\sum_{j=1}^{n} c_{ij} x_{oj} = y_i(0) \qquad (3.2.9)$$

contain all the information present in (3.2.7) provided that $y_i(0) \neq 0$. If $y_i(0)=0$ for all i, $Cx = 0$ and the following analysis may be carried out with C replacing C_o. As stated earlier, x_o may be multiplied by any non-zero scalar without altering the value of μ. (3.2.9) can be satisfied by choosing a suitable scalar multiple of any given x_o, so this condition does not contribute any useful information to the problem of maximising μ. The objective is now to find

$$\max_{x_o} \left(\frac{x_o' P(0) x_o}{x_o' S x_o} \right)$$

subject to the (m-1) homogeneous constraints

$$C_o x_o = 0 \ .$$

This is solved by using (3.2.8) to eliminate (m-1) elements of x_o and then maximising μ with respect to the remaining (n-m+1) elements. Let x_o be expressed in terms of an $(m-1)\times 1$ vector ξ and an $(n-m+1)\times 1$ vector η such that

$$x_o = X\xi + Y\eta \qquad (3.2.10)$$

where X and Y are $n\times(m-1)$ and $n\times(n-m+1)$ matrices respectively. It will also be assumed that ξ and η can be expressed as

$$\xi = Mx_o \ , \quad \eta = Nx_o, \qquad (3.2.11)$$

M and N being $(m-1)\times n$ and $n\times(n-m+1)$ matrices respectively. For this

assumption to be valid it is necessary that the partitioned matrix

$$[X\,|\,Y]$$

is non-singular, then

$$\begin{bmatrix} M \\ \hline N \end{bmatrix} = [X\,|\,Y]^{-1} .$$

Within these restrictions X and Y may be chosen to be of the form most convenient for the problem under consideration. Combining (3.2.8) and (3.2.10) gives

$$C_e X\xi + C_e Y\eta = 0 ,$$

therefore

$$\xi = -(C_e X)^{-1} C_e Y \eta \qquad (3.2.12)$$

provided that $C_e X$ is non-singular. A necessary condition for $C_e X^{-1}$ to exist is that the rows of C_e must be linearly independent as must be the columns of X. It is reasonable to assume that the rows of C are linearly independent, otherwise the system hasat least two outputs that are only different by constant scaling factors and so contribute no additional information about the state. The rows of C_e are only linear combinations of the rows of C, so they too will be linearly independent. The columns of X must be linearly independent because the condition that $[X\,|\,Y]^{-1}$ exists has already been imposed. These conditions may seem restrictive but they can always be fulfilled because choosing $x_e = \begin{bmatrix} \xi \\ \eta \end{bmatrix}$ satisfies all the restrictions on X and Y.

Substituting (2.3.12) in (2.3.10) gives

$$\begin{aligned} x_e &= (-X(C_e X)^{-1} C_e Y + Y)\eta \\ &= L\eta, \text{ say,} \qquad (3.2.13) \end{aligned}$$

where L is an $n \times (n-m+1)$ matrix. (3.2.13) and (3.2.5) together yield

$$\mu = \frac{w' L' P(0) L \eta'}{\eta' L' S L \eta'} . \qquad (3.2.14)$$

This is analogous to the original unconstrained maximisation problem, so, as in (3.2.6) one can say that the maximum value of μ is the maximum eigenvalue of

$$(L'P(0)L).(L'SL)^{-1} . \qquad (3.2.15)$$

The final consideration is whether $(L'SL)^{-1}$ exists. $L'SL$ will only be singular if there exists a non-zero η while $x_c = 0$. However, it has been assumed in (3.2.11) that $\eta = Nx_c$, so, if $x_c = 0$, η must also be zero, therefore $L'SL$ is not singular. Courant & Hilbert(1963) have shown how the range in which the value of $\max_{x_c}(\mu)$, subject to the constraints $C_c x = 0$, may be found. If the eigenvalues of $P(0)S^{-1}$ are arranged in descending order

$$\lambda_1, \lambda_2, \lambda_3 \cdots\cdots, \lambda_n$$

then, with k linear homogeneous constraints

$$\lambda_k \geq \max_{x_c} \mu \geq \lambda_{k+1} .$$

So, using the constraints must reduce μ by a non-negative amount.

It will now be shown how these results can be applied to a simple example. Consider the position control of a unit mass where the restoring force is subject to a first order lag with unit time constant. The equation governing the system is

$$\ddot{x} = \frac{-u}{1 + D}$$

where x is the position, u the controlling force and D is the operator d/dt, this may be written

$$\dddot{x} + \ddot{x} = -u .$$

We shall assume that it is possible to measure the position and the velocity of the mass, but not its acceleration. Therefore the system has two observable outputs

$$y_1 = x ,$$

$$y_2 = \dot{x} .$$

Define the state variables

$$x_1 = x\ ,\quad x_2 = \dot{x}\ ,\quad x_3 = \ddot{x}\ .$$

Hence, in the terminology of (3.2.1)

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

In the cost function let $T \to \infty$, $G = 0$, $R = 1$ and

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

that is

$$J = \int_0^\infty \left[x_1^2 + u^2\right] dt.$$

If $S = I$ then

$$\mu = \frac{x_0' P(0) x_0}{x_0' x_0} \quad . \tag{3.2.16}$$

The homogeneous constraints are derived from the equations

$$x_{0_1} = y_1(0)$$

$$x_{0_2} = y_2(0)$$

therefore

$$x_{0_1} - \alpha x_{0_2} = 0 \tag{3.2.17}$$

where

$$\alpha = x_{0_1}/x_{0_2} = y_1(0)/y_2(0)\ ,$$

so

$$C_0 = \left[1 \quad -\alpha\right] .$$

If $x_{0_2} = 0$, α cannot be defined in this way, but it is straightforward to reformulate in terms of $1/\alpha$.

The simplest choices of ξ and η are

$$\xi = x_1\ ,\quad \eta = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} .$$

In this case (3.2.13) gives

$$x_0 = \begin{bmatrix} \alpha\eta_1 \\ \eta_1 \\ \eta_2 \end{bmatrix} \quad \text{so} \quad L = \begin{bmatrix} \alpha & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

It is now straightforward, using (3.2.15), to calculate the constrained optimal feedback control $K^*(\alpha)$, for any given α such that the criterion (3.2.5) is satisfied subject to the constraints (3.2.17). The results are shown in Fig.3.2 where the elements of $K^*(\alpha)$ and the optimal value of μ are given as functions of α. As a comparison the optimal values of K and μ were found for the case in which nothing was assumed about the initial case. This involves minimising the maximum eigenvalue of P(0) with respect to K which gives rise to the feedback law

$$K = [0.42 \quad 1.62] = K^* \text{ , say.}$$

The corresponding value of μ is then 7.38 . It may be seen that using the constraints on x_0 can lead to considerable reduction in μ, however, this does not necessarily imply that the improvements in the worst case costs are of the same order of magnitude. This follows because the worst case value of μ with feedback K^* corresponds to one particular initial state which, in general, cannot be attained under the constraints (3.2.17). To carry out the comparison one has to calculate the maximum value of μ that can occur using the feedbacks K^* and $K^*(\alpha)$ where x_0 must satisfy $x_{01}/x_{02} = \alpha$. The percentage improvement in the worst case values of μ is shown in Fig.3.3 where it can be seen that up to 11% reduction is possible. When the feedback K^* is used there will be a value of α associated with the eigenvector corresponding to the maximum eigenvalue of P(0). Hence, at this value of α $K^*(\alpha) = K^*$ and there is no improvement in the worst case cost.

Fig 3.1

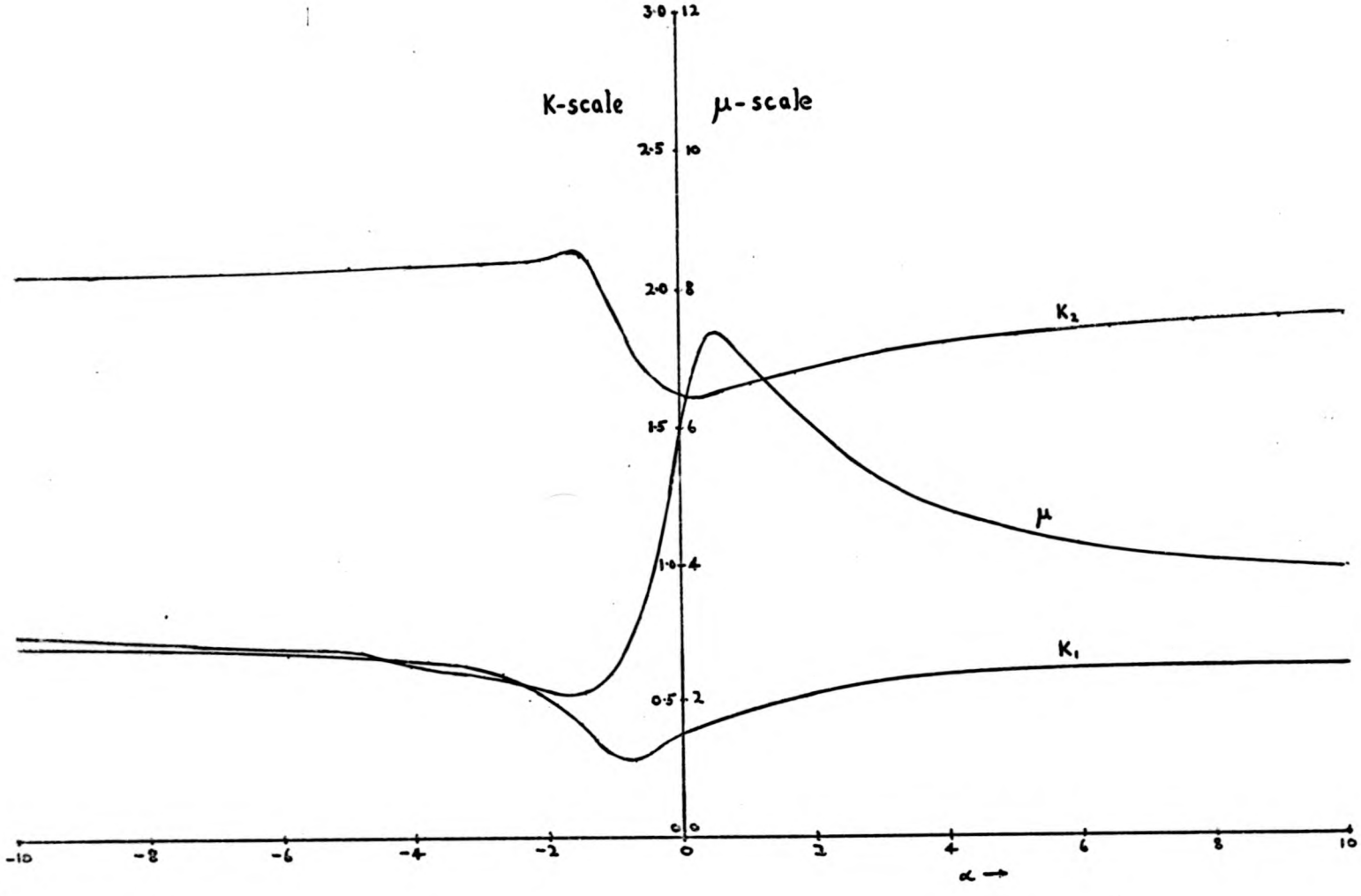
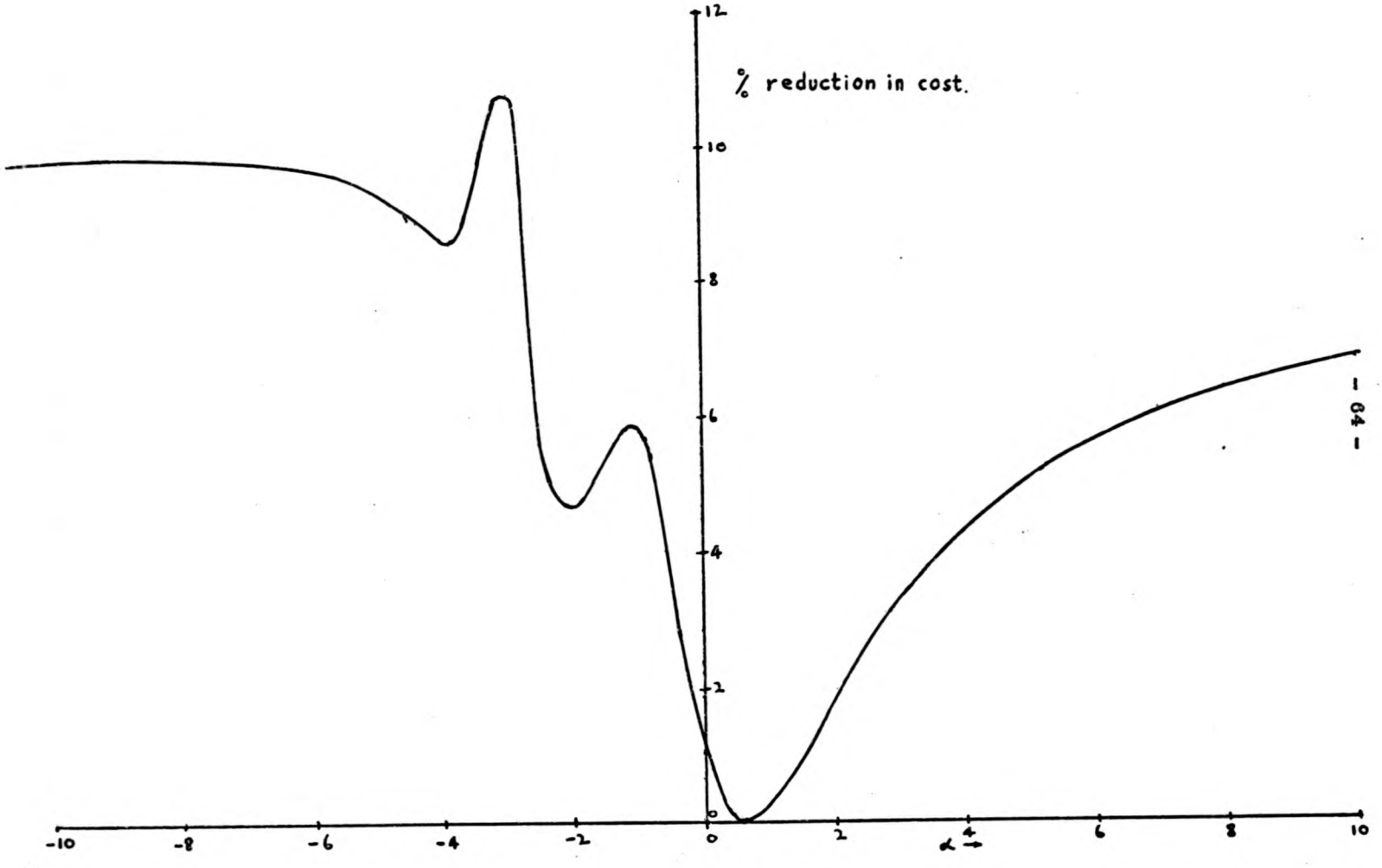

K-scale
μ-scale
3·0
12
2·5
10
2·0
8
1·5
6
1·0
4
0·5
2
0
0
K₂
μ
K₁
-10
-8
-6
-4
-2
0
2
4
6
8
10
α →

Fig. 3.2



In conclusion we may say that if one has limited knowledge of the state it is worthwhile considering using the measurement of $y(0)$ to design controllers that are better according to the criterion (3.2.4). However, the possible reduction in worst case cost must be weighed against the added complexity entailed in adjusting the elements of K according to x_0.

Section 3. A simplification of the Liapunov matrix equation.

The Liapunov matrix equation (2.3.2) plays an important part in the analysis of the linear quadratic problem as it is necessary to solve it in order to work out the cost of any control. Hence any method of simplifying its numerical calculation will be of great benfit. In the time invariant infinite time interval problem, when (2.3.2) takes the form (2.3.6), considerable reduction in the computation can be achieved if the matrix A is diagonal. This is of practical significance for a large number of distributed parameter systems with boundary control. Many partial differential operators arising from physical problems have a spectrum consisting of a countable sequence of eigenvalues, $\{\lambda_n\}$, with corresponding eigenfunctions $\{\phi_n\}$. If $Re\{\lambda_n\} \to \infty$ as $n \to \infty$ one may generally say that the components of the response associated with ϕ_n will become less significant as n increases. For this reason an approach to distributed parameter systems that can often be useful is to express the state $z(t)$ as a summation of an infinite series of the eigenfunctions, that is

$$z(t) = \sum_{i=1}^{\infty} a_i(t)\phi_i \quad ; \qquad (3.3.1)$$

the $a_i(t)$ can now be considered as the new state variables. It is then possible to truncate the series at some point suitable to give the desired accuracy and consider the problem as being of finite dimension using the $a_i(t)$ of the truncated series as state variables.

Substituting the series (3.3.1) into the original partial differential equation and forming the inner product with the eigenfunctions of the adjoint problem, $\{\phi_n^*\}$, gives the system

$$\dot{z}(t) = Az(t) + Bu(t)$$
$$z(0) = z_0 \qquad (3.3.2)$$

where A is an infinite dimensional diagonal matrix whose elements are the eigenvalues λ_n. The matrix B will depend on the form of the original problem. If the control is distributed then B will have an infinite number of rows and columns. However, if there are only a finite number, r, of control inputs which can be the case when there is control action at the boundaries, then B will only have r columns. When there is boundary control the concept of the extended definition of an operator, Brogan(1968), is useful for calculating the elments of B. This allows the boundary control to be replaced by a distributed control but the new operator B resulting will be unbounded, for example it could be an impulse function. For a detailed description of a particular system where this method is used for a diffusion equation see Parker(1970).

After truncating the series of eigenfunctions, (3.3.1) one arrives at a finite dimensional system with diagonal A. If we then wish to calculate the cost of the controller $u = KCz$ it is necessary to solve (2.3.2). Lastly, if $T \to \infty$ and the system is time invariant this equation becomes that given in (2.3.6), namely

$$P(A + BKC) + (A + BKC)'P + Q + C'K'RKC = 0. \qquad (3.3.3)$$

P is an $n \times n$ symmetric matrix which thus has $\frac{1}{2}n(n+1)$ distinct elements, hence finding P from (3.3.3) in general involves the solution of $\frac{1}{2}n(n+1)$ simultaneous linear equations. If this set of equations is to be solved on a digital computer it will be necessary to provide

$\frac{1}{2}n(n+1)\left[\frac{1}{2}n(n+1)+1\right]$ storage locations for the coefficients. As can be seen this contains a fourth power of n so a large storage facility will be needed for even quite modest n. For example, if n = 20 , the number of coefficients needed in (3.3.3) is almost 45000. It will now be shown how the number of linear equations to be solved can be reduced considerably.

Set

$$S = PB \tag{3.3.4}$$

then (3.3.3) becomes

$$PA + A'P + SKC + C'K'S' + Q + C'K'RKC = 0 . \tag{3.3.5}$$

Hence

$$p_{ij} = \frac{-d_{ij}}{\lambda_i + \lambda_j} \tag{3.3.6}$$

if $\lambda_i+\lambda_j \neq 0$. Here d_{ij} is the i,j^{th} element of

$$SKC + C'K'S' + Q + C'K'RKC .$$

Substitution for P in (3.3.4) gives the following system of equations

$$s_{ij} = \sum_{k=1}^{n} - \frac{b_{kj}}{\lambda_i + \lambda_k} \left\{ \sum_{l=1}^{r} s_{il} (KC)_{lk} + s_{kl} (KC)_{li} + q_{ik} + (C'K'RKC)_{ik} \right\} \tag{3.3.7}$$

$i = 1,2,\ldots,n$

$j = 1,2,\ldots,r.$

This reduces equation (3.3.3), which is in $\frac{1}{2}n(n+1)$ unknowns, to an equation in S which contains n*r unknowns, P is then given by (3.3.6) and if $r < \frac{1}{2}(n+1)$ the number of unknowns is reduced. If r is small and n is large, as can occur in distributed parameter systems with boundary control, the reduction can be quite considerable. If there is any particular pair of values of i and j, $\bar{i}$ and $\bar{j}$ say, for which $\lambda_{\bar{i}} + \lambda_{\bar{j}} = 0$, (3.3.5) implies that $d_{\bar{i}\bar{j}} = 0$. However r equations for $s_{\bar{i}j}$, $j=1,2,\ldots,r$, are

lost in (3.3.7). Hence if r=1 it is possible to have one case where $\lambda_i+\lambda_j=0$ as the equation in (3.3.7) that is no longer valid can be replaced by

$$d_{ij} = 0 = \sum_{k=1}^{r}\left[s_{ik}(KC)_{kj} + s_{kj}(KC)_{ki}\right] + q_{ij} + (C'K'RKC)_{ij} \qquad (3.3.8)$$

If more than one equation in (3.3.6) no longer holds true the method presented here is no longer applicable. However, if the uncontrolled system is asymptotically stable $Re[\lambda_i] < 0$ for all i hence $\lambda_i+\lambda_j \neq 0$ for any i,j. Also if the system has one zero eigenvalue and all the rest have negative real parts, that is the system is stable but not asymptotically so, the method can be applied if r=1 as then there is only one pair of eigenvalues such that $\lambda_i+\lambda_j=0$. We shall now apply the procedure described here to a distributed parameter example where it leads to a very considerable saving in computer storage requirements and processing time.

Consider the control of the temperature in a uniform bar. The control action is assumed to take place at one end of the bar and is determined by the sensed temperature at one point on the bar. The relevant equations, in dimensionless form, are

$$\frac{\partial z(t,x)}{\partial t} = \frac{\partial^2 z(t,x)}{\partial x^2} \qquad x\in[0,1]$$

subject to the boundary conditions

$$\frac{\partial z(t,1)}{\partial x} = 0$$

$$z(t,0) = u(t) = -Gz(t,a)$$

and the initial condition

$$z(0,x) = z_0(x)$$

where a is the sensing position and G is the associated positive gain. The cost funtional is chosen to be

$$J = \int_0^\infty \left[\int_0^1 z^2(t,x)dx + 0.01u^2(t) \right] dt.$$

The eigenvalues of the system are

$$= -\left\{ \frac{(2i-1)\pi}{2} \right\}^2 \qquad i = 1,2,3...$$

with corresponding eigenfunctions

$$\phi_i(x) = \phi_i^*(x) = \sin\left\{ \frac{(2i-1)\pi x}{2} \right\} .$$

Expanding $z(t,x)$ as an infinite series in $\phi_i(x)$ gives (3.3.2) where

$$A = \text{diag}(-\tfrac{1}{4}\pi^2, -\tfrac{1}{4} .9\pi^2, -\tfrac{1}{4} .25\pi^2, \ldots)$$

$$B = \text{col}(\pi, 3\pi, 5\pi, \ldots) .$$

The cost functional can now be expressed in matrix form

$$J = \int_0^\infty \left[z'Qz + u'Ru \right] dt$$

where z is now the state vector and

$$Q = \tfrac{1}{2}I$$

I being the unit matrix and

$$R = 0.01.$$

The optimisation was carried out for two different assumptions about the initial stage. First it was assumed to be constant, $z_0(x) = 1$. Secondly, no knowledge of the initial state was assumed and the gain was chosen to minimise

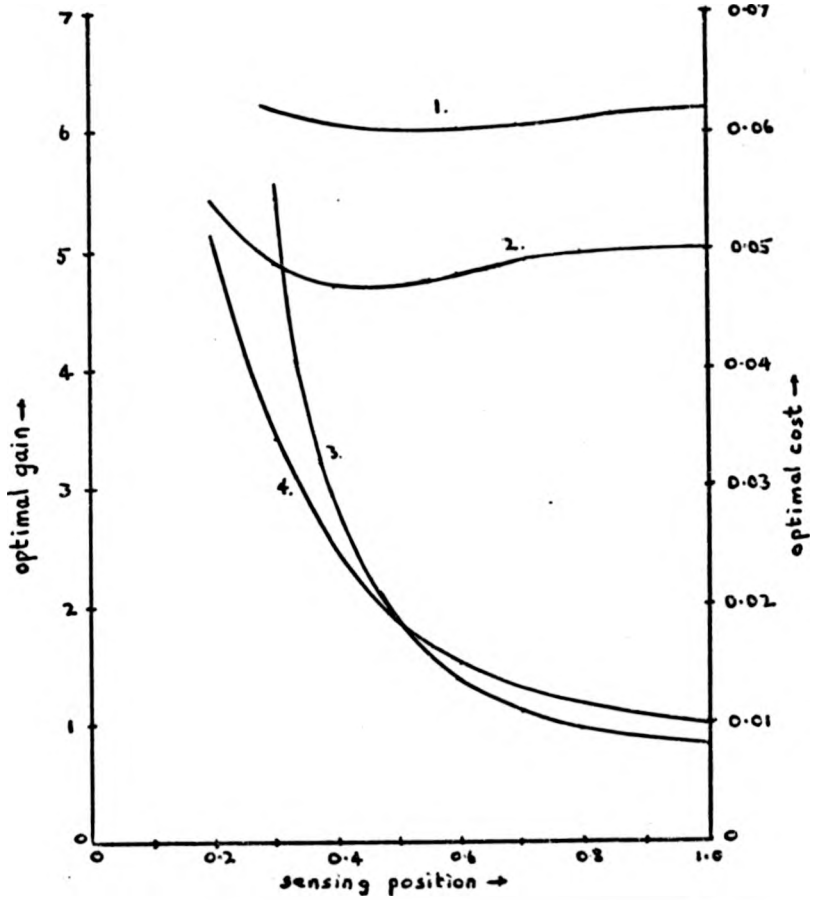
$$\max_{z_0} \frac{z_0'Pz_0}{z_0'z_0} ,$$

that is the worst case criterion (3.2.6) with $S = I$ was used. This involves minimising the maximum eigenvalue of P with respect to G. It was found necessary to take 49 eigenfunctions in the series expansion if values of the optimal cost consistent to 0.03% were to be obtained. Therefore, in the simplified method presented here, it was necessary to solve 49 simultaneous equations to calculate the matrix P. However, if the Liapunov matrix equation had

been solved directly, it would have required the solution of $\frac{1}{2}\times 49\times 50 = 1225$ equations. Over a million storage locations would be needed for the coefficients of 1225 simultaneous equations, so the direct solution of the Liapunov equation would be difficult and very expensive on present day computers.

The results of this optimisation are shown in Fig.3.3 and it can be seen how the positioning of the sensor affects the cost. It shows that if one uses the worst case criterion the position does not affect the design of the controller very much, though there is an optimum at $a \simeq 0.5$. If the initial state is known to be a constant function of x the cost is more sensitive to a with the minimum occurring at $a \simeq 0.4$. In either case it is apparent that the feedback gain G is a more important design variable, having to be considerably higher at low values of a.

Fig. 3.3.



1. Maximum eigenvalue of $P = \min_{\zeta} \max_{z_0} \left(\frac{J}{\|z_0\|^2} \right)$.

2. Cost with $z_0(x) = \text{const.}$, $\|z_0\|^2 = 1$.

3. Optimum gain corresponding to 1.

4. Optimum gain corresponding to 2.

CHAPTER 4

PROPORTIONAL-INTEGRAL-DERIVATIVE CONTROLLER DESIGN USING OPTIMAL CONTROL THEORY

Section 1. Introduction.

In the 1940's and early 50's the development of control theory was oriented towards practical systems, especially in the field of servo-mechanisms. Much of this work was concerned with the frequency domain, for example Nyquist's stability criterion, Nichols charts and Bode diagrams, and many methods were found for the "ad hoc" design of feedback systems. These are usually based on finding a control that guarantees stability and then tuning the system to give some parameter, such as gain or phase margin, a value which is known from experience to give a good response. For an account of these methods see Douce(1963) and D'Azzo & Houpis(1968). The classical control theory design procedures become more difficult to implement as the complexity of the system increases; the application to distributed parameter systems, even if it is valid, becomes very tedious as Parker(1970) has shown by applying Nyquist's stability criterion to a simple diffusion equation. Since the middle 50's control theory has shown great advances with state space analysis and optimal control playing a prominent part, however, much practical design work is carried out using classical methods. Although these have many advantages, such as the fact that they can often be carried out by hand and that practising control engineers have a great body of experience to call upon, the widespread availibility of digital computers would seem to indicate that certain systems could be analysed in more detail.

One type of controller that is common in industrial applications is the so called "proportional-integral-derivative", or PID, controller. This acts on the error signal, formed from the desired and the actual outputs, and gives a signal that is the sum of three variables. These are directly proportional to the error, its integral and its derivative; the relative sizes of the three components are then adjusted to suit the plant. One useful way of setting up these controllers is given by what are known as the Ziegler-Nicholls criteria, Ziegler & Nichols (1942). The plant is set up to give sustained oscillations and the time constants of the controller are determined from the period of the oscillation. However, this uses coefficients that are learned from experience of many installations and cannot be expected to be the best for all systems. The reasoning behind the use of PID controllers is that simple proportional controls can lead to steady state errors in the output of a system and inserting an integrator into the feedback loop removes this undesirable effect, D'Azzo & Houpis(1966). This is easily proved using Laplace transforms and is also what one would intuitively expect; the only way that the output of an integrator can be constant is for its input to be zero, hence in the steady state the error must be zero. The derivative signal is included so as to introduce an element of anticipation into the control and thus reduce overshoot.

Athans(1971) was the first to suggest an approach that might enable optimal control theory to be applied to the design of these controllers. However, he only considered a specific first order example which was presented for further discussion and development. Parker(1972) published a method of extending this work to the problem

of tracking m inputs with an n^{th} order system in such a way that there was no steady state error in any of its outputs. However, this procedure involved posing the problem in terms of the derivatives of the state and control variables, as a result it was difficult to understand the significance of the cost function. We shall present an abbreviated version of the derivation of the optimal PID controller in order to compare it with a corresponding method, given afterwards, that is not only valid for systems of infinite dimension but also does not involve the derivatives of the variables.

Section 2. The construction of an optimal PID controller with derivatives of state and control variables.

We shall consider the standard linear time invariant finite dimensional system

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$x(0) = x_0 \qquad (4.2.1)$$

$$y(t) = Cx(t) \qquad (4.2.2)$$

where x, u and y are $n\times1$, $r\times1$ and $m\times1$ vectors respectively. The objective is to design a controller such that

$$\lim_{t\to\infty} y(t) = w \qquad (4.2.3)$$

where w is a constant $m\times1$ reference vector. Using bars to indicate steady state values this implies that

$$A\bar{x} + B\bar{u} = 0 \qquad (4.2.4)$$

$$\bar{y} = C\bar{x} = w \ . \qquad (4.2.5)$$

If A is non-singular then

$$w = -CA^{-1}B\bar{u} \ . \qquad (4.2.6)$$

In order for y to reach any w it must be possible to find a $\bar{u}$ that satisfies (4.2.6), this can only be done if the dimension of u is greater or equal to that of w, in other words $r \geq m$. The full

conditions for whether it is possible to drive the system to any given state are determined by the requirements for controllability, Ogata(1967). We shall also have to define a quadratic cost function, the precise form of which will become apparant later.

Since the asymptotically stable equilibrium point of an optimally controlled system is at the origin of the state space, Ogata(1967), we must set up the state variables of this problem in such way that they are zero in the steady state. Athans(1971) and Parker(1972) achieve this by using the derivatives of $x(t)$, $u(t)$ and the error $\varepsilon(t) = w - y(t)$; in the steady state these must be zero. We shall now give an account of this approach to the problem.

The desired conditions are

$$\lim_{t\to\infty} \varepsilon(t) = 0 \qquad (4.2.7)$$

and, since it will be shown that the elements of $\dot{\varepsilon}(t)$ will be needed as state variables

$$\lim_{t\to\infty} \dot{\varepsilon}(t) = 0 \quad . \qquad (4.2.8)$$

For linear time invariant equations (4.2.8) is implied by (4.2.7). Equation (4.2.1) is now transformed so that the state vector becomes

$$\xi = \begin{bmatrix} y \\ \eta \end{bmatrix}$$

where η is an $(n-m)\times 1$ vector. Let

$$\xi = Tx$$

where T is an $n\times n$ nonsingular matrix. If we consider the partitioned form of T

$$T = \begin{bmatrix} C \\ \hdashline L \end{bmatrix} \quad ,$$

L being an $(n-m)\times n$ matrix, it can be seen that the choice of L is arbitrary except for the restriction that T must be non-singular. One possible choice for L would be so as to make

$$T = \left[\begin{array}{c|c} C_1 & C_2 \\ \hline 0 & I \end{array}\right] \qquad \text{and} \qquad \eta = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_n \end{bmatrix}$$

where C has been partitioned into the $m\times m$ and $m\times(n-m)$ matrices C_1 and C_2 and I is the $(n-m)\times(n-m)$ unit matrix. If T is to be nonsingular its determinant must be nonzero; development of det T by the bottom row gives

$$\det T = \det C_1 \ .$$

It is reasonable to assume that all the outputs of the system are linearly independent and thus C is of rank m which implies that $\det C_1 \neq 0$. Consequently $\det T \neq 0$ and T is nonsingular as desired. There are obviously other ways of defining L and one factor that will affect this choice is that the form of the controller demands feedback of $\dot{\eta}$; therefore we should like the elements of η to be those that are most easily measured. It has to be noted that $\dot{\eta}$ appears in the cost function, so different choices of L will affect the cost. However, this can usually be allowed for by varying the elements of Q in (4.2.18).

We shall now use this transformation, T, to recast (4.2.1) in terms of ξ ;

$$\dot{\xi} = TAT^{-1}\xi + TBu = F\xi + Gu \qquad (4.2.9)$$

say. Now partition F and G so that

$$F = \begin{matrix} \\ m \\ n-m \end{matrix}\!\!\begin{matrix} \overbrace{}^{m} & \overbrace{}^{n-m} \\ \end{matrix}\!\!\begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix} \qquad G = \begin{matrix} \\ m \\ n-m \end{matrix}\!\!\begin{matrix} \overbrace{}^{m} \\ \end{matrix}\!\!\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} .$$

Therefore (4.2.9) becomes

$$\dot{y} = F_1 y + F_2 \eta + G_1 u \qquad (4.2.10)$$

$$\dot{\eta} = F_3 y + F_4 \eta + G_2 u \ . \qquad (4.2.11)$$

A new set of state variables that will give rise to the necessary integral control is now defined as

$$\begin{aligned} \theta_1 &= \varepsilon \\ \theta_2 &= \dot{\varepsilon} \\ \theta_3 &= \dot{\eta} \quad ; \end{aligned} \qquad (4.2.12)$$

we shall also have to define a new control vector

$$v = \dot{u} \ . \qquad (4.2.13)$$

These choices may seem somewhat arbitrary but these vectors do satisfy the condition that they have steady state values of zero, even with variations in the system parameters. The state equations will now be formed in terms of the $(n+m)\times 1$ vector

$$\Theta = [\theta_1 \quad \theta_2 \quad \theta_3]' \ .$$

By definition, and because $\varepsilon = w - y$,

$$\begin{aligned} \dot{\theta}_1 &= \theta_2 \\ \dot{\theta}_2 &= -\frac{d^2}{dt^2}(w - y) \ ; \end{aligned} \qquad (4.2.14)$$

so, since w is being considered as a constant step input,

$$\theta_2 = -\ddot{y} \ .$$

However, from (4.2.10),

$$\ddot{y} = F_1 \dot{y} + F_2 \dot{\eta} + G_1 \dot{u} \ ,$$

therefore, since $y = -\dot{\varepsilon} = -\theta_2$,

$$\dot{\theta}_2 = F_1 \theta_2 - F_2 \theta_3 - G_1 v \ .$$

Differentiating (4.2.12) gives

$$\dot{\theta}_3 = \ddot{\eta}$$

so that, from (4.2.11-13)

$$\dot{\theta}_3 = -F_3\theta_2 + F_4\theta_3 + G_2 v \ . \qquad (4.2.16)$$

Using (4.2.14-16), the state equations can be written

$$\dot{\Theta} = \Phi\Theta + \Gamma v \qquad (4.2.17)$$

where

$$\Phi = \begin{bmatrix} 0 & I & 0 \\ 0 & F_1 & -F_2 \\ 0 & -F_3 & F_4 \end{bmatrix} \ , \quad \Gamma = \begin{bmatrix} 0 \\ -G_1 \\ G_2 \end{bmatrix} \ .$$

It has been mentioned above that the state and control variables in the cost function must tend to zero and that the time interval of integration must be infinite. The standard quadratic form is chosen for reasons explained in earlier chapters and so we must minimise

$$J = \int_0^\infty \Theta' Q \Theta + v'Rv \ dt \qquad (4.2.18)$$

where Q and R symmetric matrices of the appropriate dimension, positive semidefinite and positive definite respectively. The term $\Theta' Q \Theta$ is a measure of the deviation of the state variables from their final, zero values and Q may be chosen so that this simply becomes $\|\varepsilon\|^2$, an expression one would very much like to minimise. The term v'Rv does not explicitly penalise the control effort, but it does limit the rate at which the control u can change, something which, in practice, is desirable. It will be shown in the following section how this cost function may be interpreted in terms of deviations from steady state values rather than derivatives.

It is well known, Ogata(1967), that the solution to the optimal control problem expressed in (4.2.17) and (4.2.18) is

$$v = -R^{-1}\Gamma' P \Theta \qquad (4.2.19)$$

where P is the positive definite matrix satisfying the Riccati equation

$$P\Phi + \Phi' P - P\Gamma R^{-1}\Gamma' P + Q = 0 \ .$$

The optimal control (4.2.19) can then be written

$$v(t) = K_1\Theta_1(t) + K_2\theta_2(t) + K_3\theta_3(t) \qquad (4.2.20)$$

where K_1, K_2, and K_3 are constant matrices of the appropriate dimensions. Fig.4.1 shows the system in block diagram form and it is possible, using a standard manipulation, to remove the integrator between v and u and place it instead in every incoming branch to the summer that forms the signal v. The result of this operation is shown if Fig.4.2.

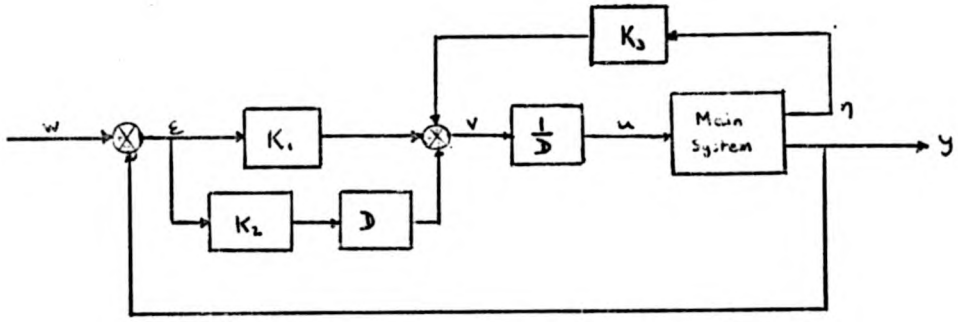
It can now be seen that u(t) is a linear combination of the error, its integral and η, that is

$$u(t) = K_1\int_0^t \varepsilon(\sigma)d\sigma + K_2\varepsilon(t) + K_3\eta(t) \; . \qquad (4.2.21)$$

In performing the block diagram manipulations it must be remembered that there is some ambiguity concerning the initial state of an integrator, this will be dealt with in the next section. $\eta(t)$ is a combination of the elements of x(t) which can be expressed in terms of y(t) and its higher derivatives, hence we have generated a controller of the PID type. If $\eta(t)$ cannot be measured completely or if one wants to restrict the feedback to only the first derivative of y(t), then one of the methods of constrained optimisation presented in chapters 2 and 3 must be used.

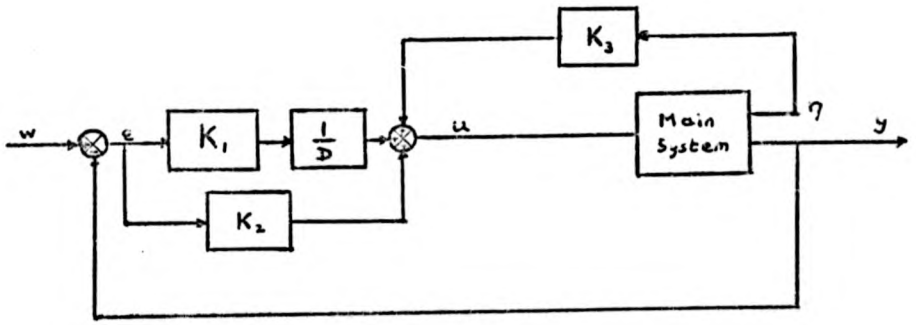
By taking Laplace transforms it is straightforward to find the steady state error in response to a ramp input of the form $w(t) = at$. As this is a class 1 system the error will be finite so there is the possibility of meeting a velocity constant specification, D'Azzo & Houpis(1968). It is useful to note that successive applications of this procedure for the design of PID controllers can be used to specify a system that will follow any prescribed polynomial input with an error that tends to zero.

Fig.4.1



$D \equiv \frac{d}{dt}$

Fig.4.2



The use of the differentials $\dot{\epsilon}$, $\dot{\eta}$ and $\dot{u}$ makes it difficult to form a rational choice for the elements of Q and R, moreover, if one wishes to deal with infinite dimensional systems there is no guarantee that these differentials exist. A method will now be presented that circumvents both these problems and yet gives the same optimal PID controller as before.

Section 3. The construction of an optimal PID controller for systems of infinite dimension.

The equations governing the system are taken to be those used in chapter 1, namely (1.1.1). If we make the added assumption that the system is time invariant these are

$$\begin{aligned} \dot{z}(t) &= Az + Bu \\ z(0) &= z_0, \end{aligned} \tag{4.3.1}$$

where z and u are elemnts of Hilbert spaces H and U repectively. In order to be able to use the procedure for deriving the optimal control given in chapter 1 the same assumptions must be made about the operators, that is A must be the infinitesimal generator of a strongly continuous semigroup T(t) and B must be a bounded linear operator. Let the output of the system be

$$y(t) = Cz(t) \tag{4.3.2}$$

where y(t) is an element of Hilbert space Y and $C:H \to Y$ is a bounded linear operator. The objective is for the output y(t) to follow a constant input $w \in Y$ with zero steady state error.

In order to frame the problem in such a way that the optimal control theory can be used it is necessary that H can be decomposed into two subspaces one of which is Y and let the other be called N, that is

$$H = Y \oplus N.$$

Kato(1966) gives the conditions for this to be possible. The implications of this assumption are that any $y \in Y$ and $\eta \in N$ define a $z \in H$ by the linear relationship

$$z = X_1 y + X_2 \eta \ . \tag{4.3.3}$$

Conversely any $z \in H$ has two components given by

$$\begin{aligned} y &= Cz \\ \eta &= Dz \end{aligned} \tag{4.3.4}$$

where $D: H \to N$ is a bounded linear operator.

Substitution of (4.3.4) into (4.3.3) gives the relationship

$$X_1 C + X_2 D = I \tag{4.3.5}$$

where I is the identity operator in H. (4.3.1), (4.3.3) and (4.3.5) may be combined to give the differential equations governing the system when the state is defined in terms of y and η.

$$\dot{y} = CAX_1 y + CAX_2 \eta + CBu$$

$$\dot{\eta} = DAX_1 y + DAX_2 \eta + DBu$$

or

$$\dot{y} = F_1 y + F_2 \eta + G_1 u \tag{4.3.6}$$

$$\dot{\eta} = F_3 y + F_4 \eta + G_2 u \tag{4.3.7}$$

provided the differentiations carried out are possible.

If the method of chapter 1 is to be used for dealing with the optimal control problem it is necessary to know the "mild" solutions of (4.3.6) and (4.3.7). The mild solution of (4.3.1) is defined as

$$z(t) = T(t-t_o)\, z(t_o) + \int_{t_o}^{t} T(t-\sigma) Bu(\sigma) d\sigma \ . \tag{4.3.8}$$

Using (4.3.5), B can be replaced by $(X_1 C + X_2 D)B$ which, together with (4.3.3) and (4.3.4) gives

$$y(t) = Cz(t) = CT(t-t_o)\left[X_1 y(t_o) + X_2 \eta(t_o)\right] + \int_{t_o}^{t} CT(t-\sigma)\left[X_1 C + X_2 D\right] Bu(\sigma)\,d\sigma$$

$$\eta(t) = Dz(t) = DT(t-t_o)\left[X_1 y(t_o) + X_2 \eta(t_o)\right] + \int_{t_o}^{t} DT(t-\sigma)\left[X_1 C + X_2 D\right] Bu(\sigma)\,d\sigma$$

or

$$y(t) = T_1(t-t_o)y(t_o) + T_2(t-t_o)\eta(t_o) + \int_{t_o}^{t}\left[T_1(t-\sigma)G_1 u(\sigma) + T_2(t-\sigma)G_2 u(\sigma)\right]d\sigma \qquad (4.3.9)$$

$$\eta(t) = T_3(t-t_o)y(t_o) + T_4(t-t_o)\eta(t_o) + \int_{t_o}^{t}\left[T_3(t-\sigma)G_1 u(\sigma) + T_4(t-\sigma)G_2 u(\sigma)\right]d\sigma \qquad (4.3.10)$$

where

$$T_1(t) = CT(t)X_1\ ,\ T_2(t) = CT(t)X_2\ ,\ T_3(t) = DT(t)X_1\ ,$$

$$T_4(t) = DT(t)X_2\ ,\ G_1 = CB\ ,\ G_2 = DB\ ,$$

these last two relationships being the same as in (4.3.6) and (4.3.7).

The objective is for the system output $y(t)$ to follow some constant input w with zero steady state error. If optimal control theory is to be used, as has been mentioned before, the state variables must tend to zero as $t \to \infty$. In section 2 the derivatives were used as they must be zero at equilibrium. However, here we shall use the differences between $y(t)$, $\eta(t)$, $u(t)$ and their steady state values $\bar{y}$, $\bar{\eta}$, and $\bar{u}$. We shall also assume that the system is such that it is possible to find a $\bar{u}$ which leads to an output $\bar{y}$ that equals w in the steady state. It is difficult to say very much about this problem for the general infinite dimensional system, each case is best considered on its own merits.

As has been pointed out before, it is a well known result of classical control theory that one needs integral control in order to

ensure zero steady state error in response to a constant input, this is also the result of section 2. Hence a new variable $\psi(t)$ is introduced where

$$\psi(t) = \psi(t_o) + \int_{t_o}^{t} \varepsilon(\sigma)d\sigma \qquad (4.3.11)$$

and the error $\varepsilon(t) = w - y(t)$. We are now in a position to define the new state variables

$$\theta_1(t) = \psi(t) - \bar{\psi} \qquad (4.3.12)$$

$$\theta_2(t) = \bar{y} - y(t) = w - y(t) = \varepsilon(t) \qquad (4.3.13)$$

$$\theta_3(t) = \eta(t) - \bar{\eta} \ . \qquad (4.3.14)$$

The mild solutions for $\theta_1(t)$, $\theta_2(t)$, and $\theta_3(t)$ must now be formed, to do this it is necessary to find the equilibrium position in terms of (4.3.9) and (4.3.10). If y and η equal $\bar{y}$ and $\bar{\eta}$ at time t_o then, if $u=\bar{u}$, they will remain unchanged for all subsequent time t. Hence

$$\bar{y} = T_1(t-t_o)\bar{y} + T_2(t-t_o)\bar{\eta} + \int_{t_o}^{t}\left[T_1(t-\sigma)G_1\bar{u} + T_2(t-\sigma)G_2\bar{u}\right]d\sigma \qquad (4.3.15)$$

$$\bar{\eta} = T_3(t-t_o)\bar{y} + T_4(t-t_o)\bar{\eta} + \int_{t_o}^{t}\left[T_3(t-\sigma)G_1\bar{u} + T_4(t-\sigma)G_2\bar{u}\right]d\sigma. \qquad (4.3.16)$$

Subtracting $\bar{\psi}$ from both sides of (4.3.11), (4.3.9) from (4.3.15) and (4.3.16) from (4.3.10) gives

$$\psi(t) - \bar{\psi} = \psi(t_o) - \bar{\psi} + \int_{t_o}^{t}\varepsilon(\sigma)d\sigma$$

$$\bar{y} - y(t) = \theta_2(t) = T_1(t-t_o).\left[\bar{y} - y(t_o)\right] - T_2(t-t_o).\left[\eta(t_o) - \bar{\eta}\right] + \int_{t_o}^{t}\left\{-T_1(t-\sigma)G_1\left[u(\sigma) - \bar{u}\right] - T_2(t-\sigma)G_2\left[u(\sigma) - \bar{u}\right]\right\}d\sigma$$

$$\eta(t) - \bar{\eta} = \theta_3(t) = -T_3(t-t_o).\left[\bar{y} - y(t_o)\right] + T_4(t-t_o).\left[\eta(t_o) - \bar{\eta}\right] + \int_{t_o}^{t}\left\{T_3(t-\sigma)G_1\left[u(\sigma) - \bar{u}\right] + T_4(t-\sigma)G_2\left[u(\sigma) - \bar{u}\right]\right\}d\sigma \ .$$

Defining

$$v(t) = u(t) - \bar{u}$$

and using the definitions of θ_1, θ_2 and θ_3 yields

$$\theta_1(t) = \theta_1(t_0) + \int_{t_0}^{t} \theta_2(\sigma)d\sigma \tag{4.3.17}$$

$$\theta_2(t) = T_1(t-t_0)\theta_2(t_0) - T_2(t-t_0)\theta_3(t_0) - \int_{t_0}^{t} \left[T_1(t-\sigma)G_1 v(\sigma) + T_2(t-\sigma)G_2 v(\sigma)\right]d\sigma \tag{4.3.18}$$

$$\theta_3(t) = -T_3(t-t_0)\theta_2(t_0) + T_4(t-t_0)\theta_3(t_0) + \int_{t_0}^{t} \left[T_3(t-\sigma)G_1 v(\sigma) + T_4(t-\sigma)G_2 v(\sigma)\right]d\sigma \tag{4.3.19}$$

(4.3.17), (4.3.18) and (4.3.19) now give mild solutions for the new state variables

$$\theta(t) \in \Theta = \begin{bmatrix} \theta_1(t) \\ \theta_2(t) \\ \theta_3(t) \end{bmatrix} = S(t-t_0)\theta(t_0) + \int_{t_0}^{t} S(t-\sigma)\Gamma v(\sigma)d\sigma \tag{4.3.20}$$

where the semigroup $S(t)$ is given by the matrix of operators

$$S(t) = \begin{bmatrix} I_Y & 0 & 0 \\ 0 & T_1(t) & -T_2(t) \\ 0 & -T_3(t) & T_4(t) \end{bmatrix}, \tag{4.3.21}$$

from which it can be shown using the definitions of T_1, T_2, T_3, T_4 with (4.3.5) that $S(t)$ satisfies (1.1.3). Also

$$\Gamma = \begin{bmatrix} 0 \\ -G_1 \\ G_2 \end{bmatrix}.$$

I_Y is the identity operator in Y.

Having shown that a mild solution exists the only other condition to be fulfilled before the method of chapter 1 can be used to solve the optimal control problem is that the operator Γ must be bounded. Since B, C and D are bounded, G_1, G_2 and hence Γ must also be bounded.

It is of interest to derive the differential equations that govern Θ_1, Θ_2 and Θ_3. Subtraction of the equilibrium conditions, obtained by putting $\dot{y}=0$ and $\dot{\eta}=0$ in (4.3.6) and (4.3.7), from (4.3.6) and (4.3.7) gives

$$\frac{d}{dt}\begin{bmatrix}\theta_1(t)\\ \theta_2(t)\\ \theta_3(t)\end{bmatrix} = \begin{bmatrix}0 & I & 0\\ 0 & F_1 & -F_2\\ 0 & -F_3 & F_4\end{bmatrix}\cdot\begin{bmatrix}\theta_1(t)\\ \theta_2(t)\\ \theta_3(t)\end{bmatrix} + \begin{bmatrix}0\\ -G_1\\ G_2\end{bmatrix} v(t) \qquad (4.3.22)$$

where it has been assumed that all the differentiation carried out is valid. (4.3.22) is exactly the same as the state equations derived by Parker(1972) except that the Θ_1, Θ_2, Θ_3 there are the differentials of those here, see (4.2.17).

To solve the optimal control problem we must first define a cost functional

$$J = \int_{t_0}^{\infty} \left[\langle \theta, Q\theta \rangle + \langle v, Rv \rangle \right] dt \qquad (4.3.23)$$

where Q and R are self adjoint operators, positive semi-definite and positive definite respectively. The cost functional has been chosen of this form because it is shown in chapter 1 that together with either (4.3.20) or (4.3.22) it yields a time invariant linear optimal feedback control

$$v(t) = K_1\theta_1(t) + K_2\theta_2(t) + K_3\theta_3(t) \qquad (4.3.24)$$

or

$$u(t) - \bar{u} = K_1\left[\psi(t) - \bar{\psi}\right] + K_2\left[w - y(t)\right] + K_3\left[\eta(t) - \bar{\eta}\right]. \qquad (4.3.25)$$

In practice it would be difficult to implement the control (4.3.25) exactly because the $\bar{u}$, $\bar{\psi}$ and $\bar{\eta}$ depend on the parameters of the system which cannot be known precisely. Consider the case when

$$u(t) = K_1\psi(t) + K_2\varepsilon(t) + K_3\eta(t) + \zeta \qquad (4.3.26)$$

where $\zeta \in U$ is some constant allowing for both the values of the steady states in (4.3.25) and for any errors in their estimates

used in implementing the control. We know in general that there will be an equilibrium position so (4.3.26) may be used to give

$$\bar{u} = K_1\bar{\psi} + K_3\bar{\eta} + \zeta \quad . \tag{4.3.27}$$

Now, subtract (4.3.27) from (4.3.26) to obtain

$$u(t) - \bar{u} = K_1\left[\psi(t) - \bar{\psi}\right] + K_2\epsilon(t) + K_3\left[\eta(t) - \bar{\eta}\right]$$

which is exactly the form of the optimal control (4.3.25). So, we may say that any value of ζ, provided that it is constant, gives an optimal feedback control. Also, since the optimal control is asymptotically stable, all the steady state values used throughout exist and in particular the limit of $\epsilon(t)$ is zero, the design objective.

The result in section 2 gives the control when ζ is zero, the most obvious choice. It has been shown that a PID controller can be designed using optimal control theory from two different viewpoints, as in sections 2 and 3, and moreover they both give the same controller; it is interesting, therefore, to look at the differences between these methods. If Q and R are the same in both approaches to the problem and ζ is chosen equal to zero one obtains identical controllers even though the θ's and v in one cost function are the derivatives of those in the other. This may seem slightly surprising but it follows from the fact that both the system and the optimal control are linear and time invariant. Differentiating the state equations and the equation giving the feedback law transform one optimal control problem into the other where state variables and controls are simply replaced by their derivatives.

Another point worthy of comment is the fact that the control is optimal regardless of the choice of ζ. The system can adjust its equilibrium position to accommodate any ζ because the output of an integrator can take on any value in the steady state, so $\bar{\psi}$ assumes

a value that takes care of the situation. Finding $\bar{\psi}$ in terms of ζ will involve solving a linear equation, but since they are both elements of U this equation will, in general, have a solution.

In conclusion we may summarize the result by saying that one can design a proportional-integral-derivative type controller using optimal control theory subject to the following conditions. The operator A is the infinitesimal generator of a strongly continuous semigroup T(t). The state z(t) can be expressed in terms of componenents $y(t)$ and $\eta(t)$ where

$$y(t) = Cz(t),\ \eta(t) = Dz(t),\ z(t) = X_1 y(t) + X_2 \eta(t)$$

and C and D are bounded operators. Finally it must be possible to find a steady state control $\bar{u}$ that drives the system to have a steady state output y equal to any prescribed $w \in Y$.

These methods for the design of PID controllers will now be demonstrated on a simple example. We shall consider the 2nd order example shown in Fig.4.3. Here it can be seen that only a simple feedback loop has been used. Taking the x_1 and x_2 shown as state variables and defining θ_1, θ_2, θ_3 as in (4.3.12)-(4.3.14) we obtain

$$\frac{d}{dt}\begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\ 0 & -2 & -1\\ 0 & 0 & -1\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\end{bmatrix} + \begin{bmatrix}0\\ 0\\ 1\end{bmatrix} v$$

The cost function is chosen as

$$J = \int_0^\infty \left[\zeta^2(t) + 0.01\, v^2(t)\right] dt\ ,$$

that is

$$Q = \mathrm{diag}(1,0,0) \quad \text{and} \quad R = 0.01.$$

Taking R this small indicates that our prime concern is minimising the integral of the error squared. Solution of the Riccati equation yields the feedback law

$$v = 10.00\theta_1 + 4.15\theta_2 - 2.05\theta_3$$

Fig 4.3

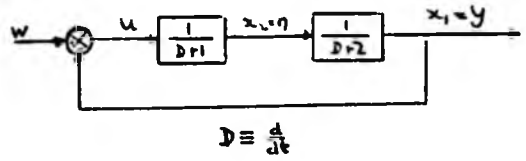


Fig 4.4

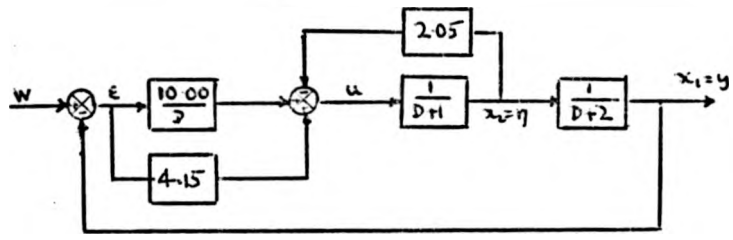
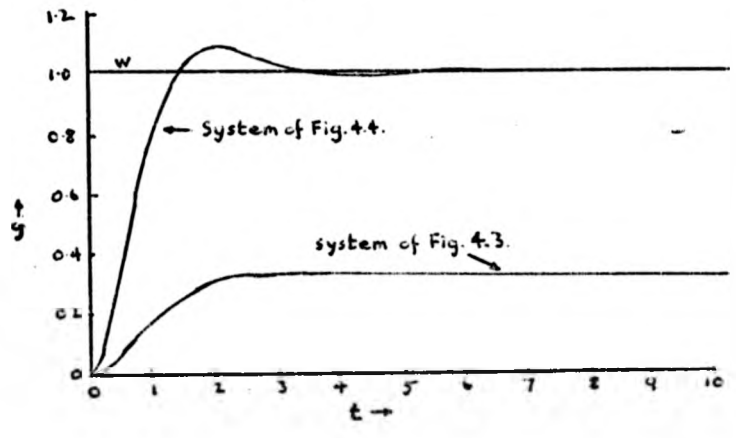


Fig. 4.5



so

$$K_1 = 10.00 \ , \quad K_2 = 4.15 \ , \quad K_3 = -2.05$$

Therefore the control system can be constructed according to Fig.4.4 and the responses of the two systems to an input $w = 1$ are shown in Fig.4.5. It can be seen that the original system has a very poor response in that there is an extremely large steady state error. The system of Fig.4.4 with the PID controller has a very satisfactory output with zero steady state error.

Also, if the input is the ramp $w = t$, the error in the first system becomes infinite, whereas that in the system of Fig.4.4 tends to 0.0166. It should be finally pointed out that in this case feedback of η is equivalent to derivative control. This follows directly from the form of the transfer functions since

$$\dot{y} + 2y = \eta \ ,$$

hence

$$\begin{aligned} u(t) &= 10\int_0^t \varepsilon(\sigma)d\sigma + 4.15\varepsilon(t) - 2.05\eta(t) + 3_1 \\ &= 10\int_0^t \varepsilon(\sigma)d\sigma + 8.25\varepsilon(t) + 2.05\dot{\varepsilon}(t) + 3_2 \end{aligned}$$

where 3_1 and 3_2 are arbitrary constants depending on the initial states of y, η and the output of the integrator.

Section 4. A distributed parameter PID controller.

In this section we shall consider the application of the preceding methods to the design of a PID controller for a distributed parameter system. The system concerns the population growth of some animal and the age distribution of its members. Since the size of the population depends on two variables, time and the ages of the members, one has to deal with partial differential equations.

Firstly we shall derive the equations governing the system and in order to do that we have to define the population age density $z(t,x)$ in a manner analogous to probability density: the number of members of the population between the ages of x and $x+dx$ at time t equals $z(t,x)dx$. By considering the balance of population entering and leaving this infinitesimal age class in the time dt it is possible to derive the partial differential equations. If the death rate at age x is $D(x)$ the number dying in time dt is $D(x)z(t,x)dxdt$. The younger members entering this class are those between the ages of $x-dt$ and x , which from the definition of $z(t,x)$ is $z(t,x)dt$, similarly the elder ones leaving amount to $z(t,x+dx)dt$. Hence the overall balance is given by

$$z(t+dt,x) - z(t,x)\,dx = z(t,x)dt - z(t,x+dx)dt - D(x)z(t,x)dxdt$$

which on dividing through by $dxdt$ gives

$$\frac{\partial z}{\partial t}(t,x) = -\frac{\partial z}{\partial x}(t,x) - D(x)z(t,x) \qquad (4.4.1.)$$

This derivation is only valid if $z(t,x)$ is differentiable but (4.4.1.) has a weak solution that is easily derived from first principles which deals with the problem of discontinuous $z(t,x)$. Consider a group of animals of some age x_0 at time t_0 and its progress in time as it grows older and dies. The number in the group is initially proportional to $z(t_0,x_0)$ and τ years later it will be proportional to $z(t_0+\tau,x_0+\tau)$. The decline in the size of this group due to the death rate is an ordinary differential equation in τ , that is

$$\frac{dz}{d\tau} = - D(x_0+\tau)z(t_0+\tau,x_0+\tau)$$

which has the solution

$$z(t_0+\tau,x_0+\tau) = z(t_0,x_0)\exp\left[\int_0^{\tau} -D(x_0+\xi)d\xi\right] \qquad (4.4.2.)$$

provided that D is integrable. There are some points worth noting about (4.4.2.) apart from that if differentiated it satisfies (4.4.1.). Firstly it shows that the solution is given by a semigroup which satisfies the conditions of Chapter 1 and that if $D(x)\to\infty$ as $x\to x_{max}$ then the population cannot have any members older than x_{max} . Also it represents a travelling wave that is being continuously attenuated by the exponential function as it moves along.

The next step is to fit these equations in with the birth rate. This is simplified if we only consider one sex as then a single self-contained equation results. If both sexes are taken into account and different birth and death rates for males and females allowed for, then two coupled partial differential equations result. Since

maternity is much more easily observed and measured than paternity it is usual only to deal with the female population. The age specific fecundity rate, $f(x)$, is defined as the proportional rate of giving birth to female offspring by mothers of age x. Thus the total rate of female births is given by

$$\int_0^\infty f(x)z(t,x)dx \quad .$$

Since animals over a certain age never give birth the upper limit of the integral can be made finite in practice. If we now observe that the number of animals between the ages of 0 and dt is both equal to $z(t,0)dt$ and to the number born in the last dt time units we obtain

$$z(t,0)dt = dt \int_0^\infty f(x)z(t,x)dx \quad .$$

Thus we have defined the system completely by

$$\frac{\partial z}{\partial t}(t,x) = -\frac{\partial}{\partial x}z(t,x) - D(x)z(t,x) \qquad (4.4.3.)$$

subject to the boundary condition

$$z(t,0) = \int_0^\infty f(x)z(t,x)dx \qquad (4.4.4.)$$

provided that we also have the initial condition

$$z(0,x) = z_0(x) \quad . \qquad (4.4.5.)$$

It can be seen that a discontinuity will arise if

$$z_0(0) \neq \int_0^\infty f(x)z_0(x)dx$$

but this can be dealt with by means of the weak solution (4.4.2.).

The best way of ascertaining whether the system will grow and at what rate is to find the eigenvalues of the equations (4.4.3.) and (4.4.4.). This can be achieved by finding the λ and $\phi(x)$ such that $z(t,x) = e^{\lambda t}\phi(x)$ satisfies the equations. On substitution one obtains

$$\frac{d\phi}{dx}(x) + \left[\lambda + D(x)\right]\phi(x) = 0 \qquad (4.4.6.)$$

and $$\phi(0) = \int_0^\infty f(x)\phi(x)dx \quad . \qquad (4.4.7.)$$

(4.4.6.) has the solution

$$\phi(x) = \phi(0)\exp\left[-\int_0^x \lambda + D(\xi)d\xi\right]$$

so from (4.4.7.)

$$1 = \int_0^\infty e^{-\lambda x} f(x)\exp\left[-\int_0^x D(\xi)d\xi\right]dx \quad . \qquad (4.4.8.)$$

$f(x)$ and the exponential function are always positive in the range of integration so the right hand side is a monotonically decreasing function of λ. Moreover it tends to $+\infty$ as $\lambda \to -\infty$ and to zero as $\lambda \to +\infty$ therefore (4.4.8.) must have one, and only one, real solution for λ. Also it follows that whether the population declines, is stationary or grows depends on whether $\int_0^\infty f(x)\exp\left[-\int_0^x D(\xi)d\xi\right]dx$ is less than, equal to, or greater than one respectively. Obviously in the real world a population cannot grow indefinitely, there are limits which have not been taken into account. The most important is that $D(x)$ depends on z when the population is so large that there is not enough food to go round, it is also possible that fertility is reduced by overcrowding. However, we shall only consider the case where the animal numbers are not

large enough to be impinging on such environmental limits.

Having looked at the general properties we shall consider the specific problem of finding an optimal culling policy. This is quite a common problem that arises when an animal population is growing faster than is desired for some ecological management scheme. The most frequent cases are those in which there is conflict between some species and a human economic activity. For example seals that interfere with fisheries and Oyster Catchers which consume large numbers of cockles. It is not claimed that the methods presented here are going to give a superior scheme for practical management, especially for animals that have one breeding season per year and are subject to an annual cull. In this instance a good case can be made out for using difference equations, Leslie(1945). However, it should be possible to extend the PID design methods to difference equations and then apply it to this sort of problem; I believe that this would be a fruitful source of future work. Despite the limitations many useful points turn up which show the potential of this PID design procedure, and its limitations.

The problem is to drive the population age density to its desired value $z_d(x)$ by culling $u(t,x)$ per unit time of those animals aged x . Equation (4.4.3.) now has to be modified thus

$$\frac{\partial z}{\partial t} = -\frac{\partial z}{\partial x} - Dz - u \quad . \qquad (4.4.9.)$$

One's immediate response to this problem is to try simple proportional control on the error $z_d - z$ such that

$$u = K(z - z_d) \qquad (4.4.10.)$$

where K is an operator that maps variables on the interval $[0,\infty)$ onto that same interval. The drawback to this approach is that the form of $z_d(x)$ is limited; this can be seen by putting $z=z_d$ in (4.4.9.) and (4.4.10.) together with the equilibrium condition $\partial z/\partial t = 0$. Then

$$\frac{dz_d}{dx} + D(x)z_d = 0 \qquad (4.4.11.)$$

so $z_d(x)$ has to be proportional to $\exp\left[-\int_0^x D(\xi)d\xi\right]$. This is a very restrictive condition and is certainly undesirable in the face of perturbations in $D(x)$ which could well occur, for example, during a hard winter.

Therefore the use of some form of integral control would seem to be indicated; this has the great advantage of always driving the system to a state in which there is no steady state error even when there are variations in the system parameters. Obviously there are limits on how big these variations can be, for instance they could not be so large as to alter the system from being stable to being unstable.

We now proceed to analyse the problem using the methods of this chapter. Firstly the error is defined as

$$\varepsilon(t,x) = z_d(x) - z(t,x)$$

and the integral of the error, $\psi(t,x)$ is given by

$$\frac{\partial \psi}{\partial t}(t,x) = \varepsilon(t,x) \quad .$$

Using the notation of section 3 it must first be noted that since $y = z \quad \eta$ disappears from the equations so

$$\begin{aligned} \theta_1 &= \psi - \bar{\psi} \\ \theta_2 &= \varepsilon \\ v &= u - \bar{u} \end{aligned} \qquad (4.4.12.)$$

and thus

$$\frac{\partial}{\partial t}\begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix} = \begin{bmatrix}0 & 1 \\ 0 & -\frac{\partial}{\partial x} - D(x)\end{bmatrix}\begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix} + \begin{bmatrix}0 \\ 1\end{bmatrix} v \quad . \qquad (4.4.13.)$$

A cost function has to be defined and, as always, this is one of the most difficult problems when applying optimal control in practical situations. Obviously the error must be driven to zero as quickly as possible but since an actual culling operation in the field will be expensive some compromise has to be found between low errors and the number of kills. Therefore the cost function will be taken as

$$J = \int_0^\infty \langle \theta_1, Q_1 \theta_1 \rangle + \langle \theta_2, Q_2 \theta_2 \rangle + \langle v, Rv \rangle \, dt \qquad (4.4.14.)$$

where the inner product $\langle \alpha, \beta \rangle$ is defined as

$$\int_0^\infty \alpha(x)\beta(x)\,dx \quad .$$

It can be seen apart from the error θ_2 and the control v there is also a term in θ_1 , the deviation of the error integral from its equilibrium, we shall now show why

this is necessary to be consistent with the setting of the problem. The strong form of the Riccati equation (1.2.2.) is

$$PA + A^*P - PBR^{-1}B^*P + Q = 0$$

and in partitioned form for the system under consideration becomes

$$\begin{bmatrix} P_1 & P_2 \\ P_2^* & P_3 \end{bmatrix}\begin{bmatrix} 0 & I \\ 0 & F \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ I & F^* \end{bmatrix}\begin{bmatrix} P_1 & P_2 \\ P_2^* & P_3 \end{bmatrix} - \begin{bmatrix} P_1 & P_2 \\ P_2^* & P_3 \end{bmatrix}\begin{bmatrix} 0 \\ I \end{bmatrix} R^{-1}\begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} P_1 & P_2 \\ P_2^* & P_3 \end{bmatrix} + \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} = 0 \qquad (4.4.15.)$$

where $F = \frac{-\partial}{\partial x} - D(x)$ and P_1, P_3, R, Q_1, Q_2 are self-adjoint operators. We shall now proceed formally and expand (4.4.15.) to obtain three equations in

$$\begin{aligned} &P_2R^{-1}P_2^* + Q_1 = 0 \\ &P_1 + P_2F - P_2R^{-1}P_3 = 0 \qquad (4.4.16.) \\ &P_2^* + P_3F + P_2 + F^*P_3 - P_3R^{-1}P_3 + Q = 0 \end{aligned}$$

If Q_1 , is chosen to be identically zero then from the first of these three equations and by inspection of the second two we obtain the solutions

$$P_2 = 0 \quad , \quad P_1 = 0$$

with P_3 given by

$$P_3F + F^*P_3 - P_3R^{-1}P_3 + Q_2 = 0 \quad .$$

If the optimal control is given by

$$v = K_1\theta_1 + K_2\theta_2$$

then $\begin{bmatrix} K_1 & K_2 \end{bmatrix} = -R^{-1}\begin{bmatrix} 0 & I \end{bmatrix}\begin{bmatrix} P_1 & P_2 \\ P_2^* & P_3 \end{bmatrix} = \begin{bmatrix} -R^{-1}P_2^* & -R^{-1}P_3 \end{bmatrix}$.

Thus if $Q_1 = 0$ and $P_2 = 0$, K_1 is also zero with the result that no integral control is used at all. Therefore choosing $Q_1 = 0$ completely negates the purpose of formulating the problem in this way, we are then left with a simple proportional control problem with the third equation of (4.4.16.) being its Riccati equation. In this case, though the system would be stable, in general its steady state will not be the desired one so the integral of the error would increase indefinitely; this does not matter if there is no penalisation of this term in the cost function so, since we want the error to be exactly zero, Q_1, must give rise to a finite increase in the cost.

It is now possible to proceed to the derivation of the optimal control by the method of Chapter 1 using the weak semigroup form (4.4.2.) and (4.4.4.). However, in order to carry out any computation some form of discretisation has to be used and the most straightforward way is to divide the population into a finite number of age classes and solve the optimal control problem in terms of the resulting ordinary differential equations.

Let there be n age groups of width h years and let the number of members between ages ih and $(i+1)h$ be w_i. We can integrate (4.4.9.) from ih to $(i+1)h$ with respect to x

$$\int_{ih}^{(i+1)h} \frac{\partial z}{\partial t}(t,x)\,dx = -\int_{ih}^{(i+1)h} \frac{\partial z}{\partial x}(t,x)\,dx - \int_{ih}^{(i+1)h} D(x)z(t,x)\,dx - \int_{ih}^{(i+1)h} u(t,x)\,dx \quad .$$

Now, $w_i = \int_{ih}^{(i+1)h} z(t,x)\,dx$ so this becomes

$$\frac{dw_i}{dt} = -z(t,(i+1)h)+z(t,ih)-\bar{D}_i w_i-\bar{u}_i(t) \qquad (4.4.17.)$$

where $\bar{D}_i$ is the average death rate over the interval of integration with a weighting factor $z(t,x)$. We do not know beforehand what this factor will be but if the interval is small we can assume it is approximately constant so

$$\bar{D}_i = \frac{1}{h}\int_{ih}^{(i+1)h} D(x)dx \quad . \qquad (4.4.18.)$$

$\bar{u}_i(t)$ is defined as the removal rate from the i^{th} class which is given by

$$\bar{u}_i(t) = \int_{ih}^{(i+1)h} u(t,x)dx \quad . \qquad (4.4.19.)$$

All that is then needed is to express $z(t,ih)$ and $z(t,(i+1)h)$ in terms of the w_i's. The average age density in the i^{th} class is w_i/h and $z(t,ih)$ is the density at the point between the $(i-1)^{th}$ and i^{th} class; therefore we may take the mean of these two averages to give

$$z(t,ih) = \frac{1}{h}(w_i - w_{i-1})$$

hence (4.4.17.) may be written

$$\frac{dw_i}{dt} = \frac{1}{2h}(w_{i-1} - w_{i+1}) - \bar{D}_i w - \bar{u}_i \quad . \qquad (4.4.20.)$$

At the extreme ends $i = n$ and $i = 0$ other conditions hold and (4.4.20.) has to be modified. When $i = n$, w_{i+1} is not defined so it has to be estimated by linear extrapolation, that is $w_{n+1} \approx 2w_n - w_{n-1}$, thus

$$\dot{w}_n \approx -(1/h + \bar{D}_n)w_n + w_{n-1}/h - \bar{u}_n . \qquad (4.4.21.)$$

The final step in forming the finite dimensional approximation is to bring in the boundary condition (4.4.4.). The number of animals in w_0 is simply the number born in the last h years; the very young do have a finite probability of dying so the fecundity rates should be modified to allow for those that do not reach the age of h. Thus if $\bar{f}_i$ is the modified average fecundity rate over the range ih to $(i+1)h$ then

$$w_0 = h\sum_{i=0}^{n} \bar{f}_i w_i$$

or

$$w_0 = \frac{h\sum_{i=1}^{n} \bar{f}_i w_i}{1 + \bar{f}_0 h} \qquad (4.4.22.)$$

If h is less than the age of sexual maturity, which is very probable, then $\bar{f}_0 = 0$. Putting $i = 1$ in (4.4.21.) gives

$$\dot{w}_1 = \frac{1}{2h}(w_0 - w_2) - \bar{D}_1 w_1 - \bar{u}_1$$

$$= \left[\frac{\bar{f}_1}{2(1+\bar{f}_0 h)} - D_1\right] w_1 + \left[\frac{\bar{f}_2}{2(1+\bar{f}_0 h)} - \frac{1}{2h}\right] w_2 + \frac{\sum_{i=3}^{n} \bar{f}_i w_i}{2(1+\bar{f}_0 h)} - \bar{u}_1 , \qquad (4.4.23.)$$

(4.4.20.), (4.4.21.) and (4.4.23.) may now be combined to give the full matrix equations

$$\begin{bmatrix} \dot{w}_1 \\ \dot{w}_2 \\ \dot{w}_3 \\ \vdots \\ \dot{w}_n \end{bmatrix} = \begin{bmatrix} \bar{f}_1/\alpha - \bar{D}_1 & \bar{f}_2/\alpha - 1/2h & \bar{f}_3/\alpha & \bar{f}_4/\alpha & \cdots & \bar{f}_n/\alpha \\ 1/2h & -\bar{D}_2 & -1/2h & 0 & \cdots & 0 \\ 0 & 1/2h & -\bar{D}_3 & -1/2h & & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & 0 & 1/h & -(\frac{1}{h} + \bar{D}_n) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} - \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}$$

(4.4.24.)

where, for ease of notation, $\alpha = 2(1 + \bar{f}_o h)$.

In order to apply the optimal control to the finite dimensional representation of the system it is necessary to express the cost function in vector-matrix terms. We shall choose (4.4.14.) to be specifically of the form

$$J = \int_0^{\infty}\int_0^{\infty} q_1\theta_1^2(t,x) + \theta_2^2 + rv^2(t,x)\,dx\,dt \qquad (4.4.25.)$$

where q_1, and r are scalar weighting factors open to choice. Set the desired values of w to be

$$w_{di} = \int_{ih}^{(i+1)h} z_d(x)\,dx$$

then the error

$$\theta_{2i} = w_{di} - w_i .$$

and the integral of the error, in terms of the deviation from its mean, θ_{1i}, is given by

$$\dot{\theta}_{1i} = \theta_{2i}$$

Euler's (or rectangular) approximation to the spatial integrals in (4.4.25.) gives

$$J = h\int_0^\infty q_1 \sum_{i=0}^{n} \theta_{1i}^2 + \sum_{i=0}^{n} \theta_{2i}^2 + r\sum_{i=0}^{n} v_i^2 \, dt. \qquad (4.4.26.)$$

However, the variables in the interval 0 to h, with subscript zero, are not independent of the other state variables, with (4.4.22.) providing the connection. It could be possible to allow for this in the cost function but that is really an unnecessary complication. As mentioned before there are no hard and fast ways of arriving at a suitable cost function and its form is nearly always considerably determined by convenience. Therefore it is quite justified not to take into account the terms affecting the cost function over the age range 0 to h and so, leaving out the constant multiplying term h from (4.4.26.) we shall use the cost function

$$J = \int_0^\infty q_1 \sum_{i=1}^{n} \theta_{1i}^2 + \sum_{i=1}^{n} \theta_{2i}^2 + r\sum_{i=1}^{n} v_i^2 \, dt$$

or, in matrix terms

$$J = \int_0^\infty q_1 \theta_1' I \theta_1 + \theta_2' I \theta_2 + rv'iv \, dt \qquad (4.4.27.)$$

If (4.4.24.) is written

$$\dot{w} = Aw - I\bar{u}$$

the state equations (4.3.22.) which are needed for deriving the PID controller become

$$\frac{d}{dt}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} = \begin{bmatrix}0 & I\\ 0 & A\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} + \begin{bmatrix}0\\ I\end{bmatrix}v \quad ; \qquad (4.4.28.)$$

the corresponding cost function is then written

$$J = \int_0^\infty [\theta_1\ \theta_2]\begin{bmatrix}q_1 I & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} + rv'Iv\ dt\ . \qquad (4.4.29.)$$

We shall now apply this formulation to a situation for which we have real data. Beddington and Taylor (1973) give the age specific death and fecundity rates for female Red Deer (Cervus elaphus L) in Scotland, which are shown in Table 4.1. It has to be admitted that the population growth of this species might well be more appropriately treated by finite difference equations on account of its only having one breeding season per year. However, this example is more intended as an instructive illustration rather than an accurate simulation. Since the objective of all optimal control theory must ultimately be to improve methods of controlling real world systems, any step from pure abstraction towards reality will almost certainly teach us something of use upon the way.

We have to decide, in this case, how many age classes of what width must be used. Here we come up against limits of computing facilities, either core storage space or cost limits set by central processing unit time. Using the methods of Chapter 1 we have at every iteration step to solve a linear equation for the symmetrix matrix P ; for

TABLE 4.1

Age specific data for Red Deer

age	death rate	fecundity
1	0.093	0.000
2	0.013	0.000
3	0.008	0.266
4	0.010	0.282
5	0.013	0.338
6	0.008	0.380
7	0.047	0.383
8	0.058	0.391
9	0.120	0.339
10	0.281	0.237
11	0.270	0.164
12	0.332	0.170
13	0.332	0.169
14	0.332	0.169
15	0.332	0.169
16	0.332	0.169
17	0.332	0.169
18		0.169

an N^{th} order system this has $\frac{1}{2}N(N+1)$ independent elements so the storage requirements are at least $\frac{1}{2}N(N+1)(\frac{1}{2}N(N+1)+1)$. If we wished to use all the data available and to take into account 18 separate age classes we would, because of the integral control, have a 36th order system; this would necessitate over 444K of store, beyond any single computer commonly available. Therefore some limitation has to be imposed and so we shall go to the other extreme and only consider hinds up to the age of 8 which will be split into 4 2-year age classes, a fairly drastic simplification but adequate for illustrating the methods. Since the numbers in the first age group are given by those in the other 3 we only have to deal with a 6th order system when the integral control is taken account of.

For this system (4.4.24.) becomes

$$\begin{bmatrix} \dot{w}_1 \\ \dot{w}_2 \\ \dot{w}_3 \end{bmatrix} = \begin{bmatrix} 0.114 & -0.091 & 0.173 \\ 0.25 & -0.011 & -0.25 \\ 0 & 0.5 & -0.553 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} - \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

where the fact that fawns only have a probability of $0.907 \times 0.987 = 0.895$ of surviving the first two years of their lives has been allowed for by reducing the fecundity rates by this factor. In the cost function (4.4.29.) q_1 will be chosen as 0.1 as this only has to be non-zero; r will be taken to be 10 which gives high weighting to the cost of control and tacitly assumes that the deviation of the population from the desired is not critical and it is only a general aim to keep it low. Starting from a feedback gain matrix of

$$\begin{bmatrix} -0.01 & 0 & 0 & -0.5 & 0 & 0 \\ 0 & -0.01 & 0 & 0 & -0.5 & 0 \\ 0 & 0 & -0.01 & 0 & 0 & -0.5 \end{bmatrix}$$

the methods of Chapter 1 led to the optimal matrix

$$\begin{bmatrix} -0.096 & -0.027 & 0.0023 & -0.70 & -0.095 & -0.081 \\ 0.022 & -0.083 & -0.051 & -0.095 & -0.48 & 0.0031 \\ -0.016 & 0.048 & -0.086 & -0.081 & 0.0031 & -0.22 \end{bmatrix}$$

only 7 iterations were necessary to get 2% convergence in every element. On inspection it can be seen that if this matrix is partitioned into two 3 x 3 matrices, thus separating the integral and proportional controls, the diagonal terms dominate. This means that the culling rate of any particular age group predominantly depends on the error, and its integral, in that class.

The system was then simulated for certain assumptions about the desired population and its initial state. It is unrealistic for culling rates to be negative which implies releasing animals of certain ages into the population; apart from necessitating the maintenance of a stock of animals of suitable ages it would be unlikely to be a sound ecological management policy as the introduced members would probably not be accepted by the wild population. Therefore the natural death rate sets an upper limit on possible equilibrium age profiles of the population, the number in any age class determines the maximum number in all classes of greater age. This condition may be written symbolically for the continuous case as

$$z_d(x+y) \leqslant z_d(x)\exp\left[\int_x^y -D(\xi)\,d\xi\right]$$

for all x and $y \geqslant x$, with an analogous expression for the discretised system. In our example we choose as the desired population

$$w_d = [8 \quad 4 \quad 2]'$$

which satisfies the above condition. Since the population would naturally grow a continuous cull will be necessary at equilibrium and it is simple to calculate from (4.2.4.) that the necessary rates of removal from the three age classes are

$$0.894, \quad 1.456, \quad 0.894 \quad .$$

The vector v represents deviations from these figures so, in order to preserve the realistic condition of non-negative culls, the three elements of v must never be less than

$$-0.894, \quad -1.456, \quad -0.894 \text{ respectively.}$$

Figs. 4.6 and 4.7 show the results of a simulation when the initial value of w is

$$[18 \qquad 13 \qquad 10]'$$

that is the initial error vector equals

$$[10 \qquad 9 \qquad 8]'.$$

It is also assumed that at the start of the run the integral of the error is zero for all age classes. It can be seen

that in fact the optimal control never calls for a negative cull with this particular set of initial conditions. Nor, for that matter, does the simulation run into the physically meaningless situation of a negative population, though this is quite feasible when the culling rate in any one age group depends on the errors and their integrals in all classes. If either of these unrealistic situations did arise in practice it would be necessary to abandon the theoretical linear optimal control and impose hard limits on some of the variables. For example, if a negative cull rate was indicated for an age group, cropping of that age group would have to be stopped; since the population tends naturally to grow, culling would not be necessary otherwise, under such conditions one would still have negative feedback, though not acting as strongly as in the theoretically optimal case. If the optimal control tried to remove animals from an age class that was already zero again culling would have to be stopped, but as long as the desired number of members was positive the ageing process would fill the empty group and thus reduce the error, a negative feedback effect though again not optimal quantitatively. If these bounds were put into the formulation of the optimal control problem it would no longer be linear and the preceding methods invalid; the alternative would be to use Pontryagin's maximum principle for the discretised system and, as pointed out in earlier chapters, this is very unattractive computationally. However, none of these problems arises

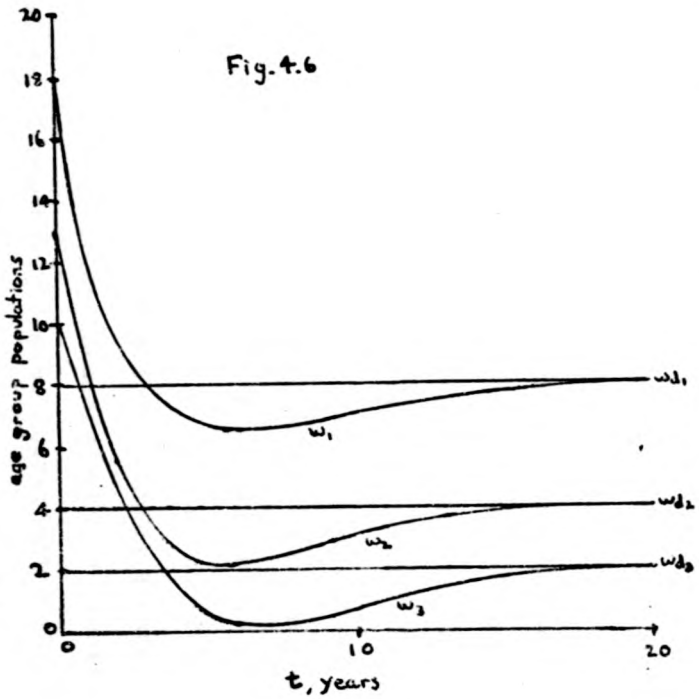

20
18
16
14
12
10
8
6
4
2
0
age group populations
0
10
20
t, years
w_1
w_2
w_3
w_{d_1}
w_{d_2}
w_{d_3}

Fig. 4.6

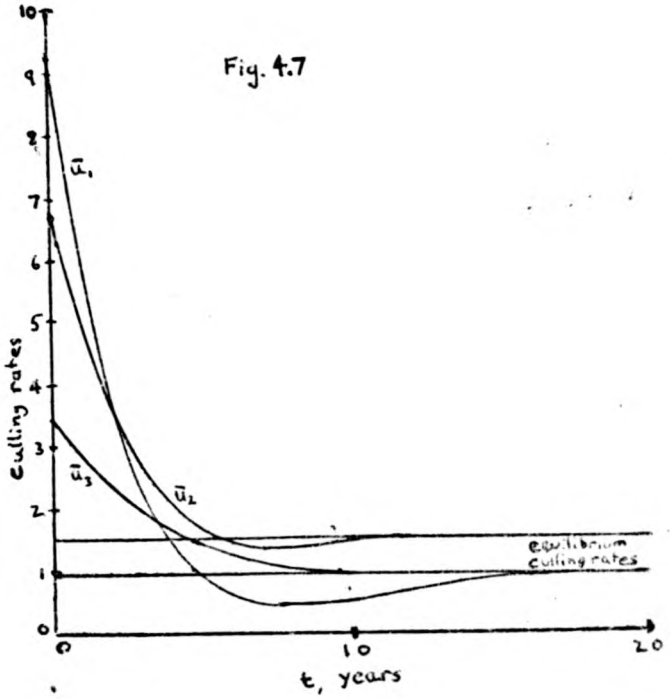

10
9
8
7
6
5
4
3
2
1
0
culling rates
0
10
20
t, years
F_1
F_2
F_3
equilibrium
culling rates

Fig. 4.7

in our simulated example and the results are very satisfactory, there is no zero steady state error and the initial response is very quick though perhaps there is more overshoot than one would intuitively like to see.

In conclusion it can be said that within rather limited terms of reference the methods of this chapter have successfully been used to design a PID controller for a practical distributed parameter system. Inevitably storage and computing time limits show themselves, an infinite dimensional system requires an infinite amount of both on a digital computer. If the methods of Chapter 1 are to be used on a typical computer with 100K of central memory we cannot solve the resulting Liapunov matrix equation for a system of greater than 24th order. Therefore we are restricted to 12th order systems when trying to control the state of the system with a PID controller. If only N outputs are being controlled then the integral control could be found for a system of order (24-N). It is possible, though, that with skilful use of mass storage devices, magnetic disc and tape, that these limits could be extended. There might well be systems, for instance in chemical processing, where very precise control is necessary, then careful consideration must be given to the discretisation scheme in order to obtain the greatest accuracy with the least number of state variables. However, considering the problems in choosing the parameters of the cost function, it is

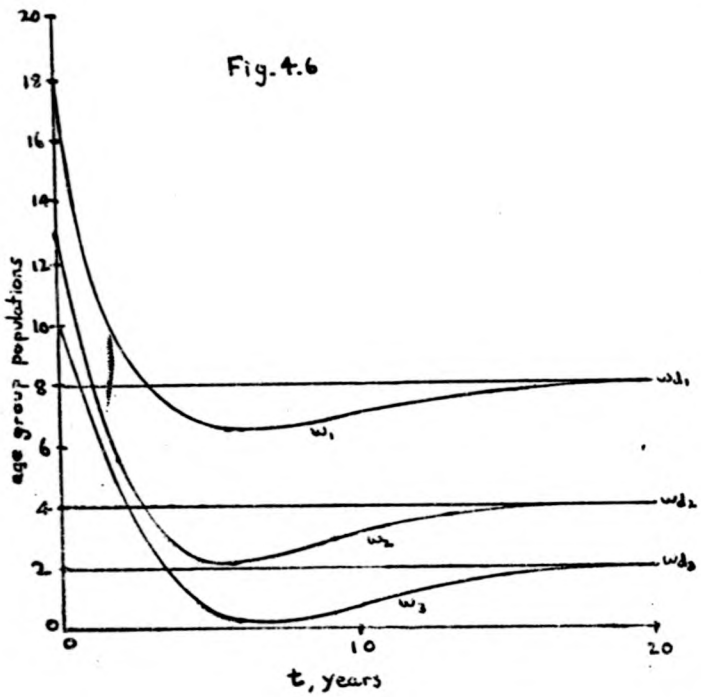

20
18
16
14
12
10
8
6
4
2
0
age group populations
0
10
20
t, years
w_1
w_2
w_3
wd_1
wd_2
wd_3

Fig. 4.6

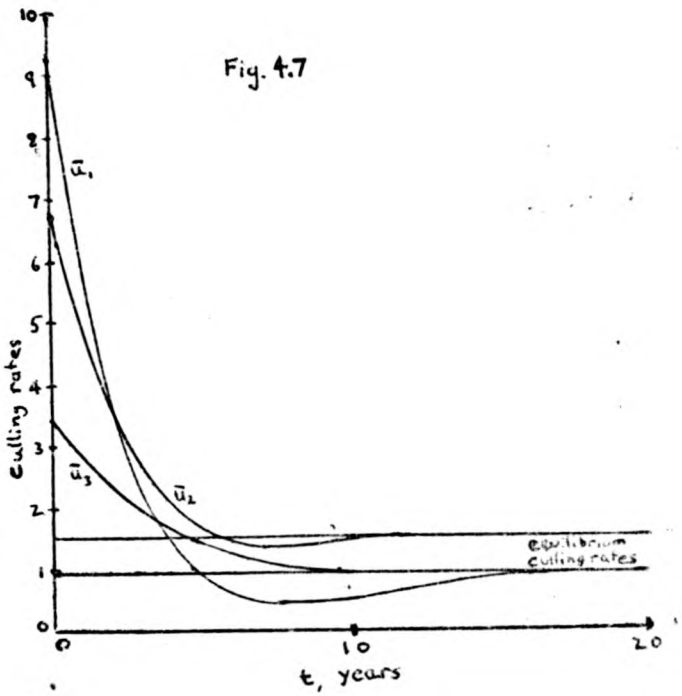

10
9
8
7
6
5
4
3
2
1
0
culling rates
0
10
20
t, years
$\bar{u}_1$
$\bar{u}_2$
$\bar{u}_3$
equilibrium
culling rates

Fig. 4.7

PAGE

MISSING

in our simulated example and the results are very satisfactory, there is no zero steady state error and the initial response is very quick though perhaps there is more overshoot than one would intuitively like to see.

In conclusion it can be said that within rather limited terms of reference the methods of this chapter have successfully been used to design a PID controller for a practical distributed parameter system. Inevitably storage and computing time limits show themselves, an infinite dimensional system requires an infinite amount of both on a digital computer. If the methods of Chapter 1 are to be used on a typical computer with 100K of central memory we cannot solve the resulting Liapunov matrix equation for a system of greater than 24th order. Therefore we are restricted to 12th order systems when trying to control the state of the system with a PID controller. If only N outputs are being controlled then the integral control could be found for a system of order (24-N). It is possible, though, that with skilful use of mass storage devices, magnetic disc and tape, that these limits could be extended. There might well be systems, for instance in chemical processing, where very precise control is necessary, then careful consideration must be given to the discretisation scheme in order to obtain the greatest accuracy with the least number of state variables. However, considering the problems in choosing the parameters of the cost function, it is

questionable whether it is worthwhile expending a considerable amount of effort on obtaining extremely accurate results. All in all, though, from the work carried out here it seems that design of optimal PID controllers for distributed parameter systems, albeit with limited accuracy, is quite feasible.

CHAPTER 5

AN ASSESSMENT OF THE METHODS: AN EXAMPLE

Section 1. Introduction.

This thesis so far has been predominantly concerned with the theory of optimal control, for both fully and partially observed systems. However, it must be remembered that the ultimate aim is to build better controllers for physical systems. It is possible to discuss in an abstract manner the limitations of the theory, for example the problems in choosing the parameters of the cost function, but the most instructive procedure is to work through a control design problem using the methods mentioned in the earlier chapters. In order to do this it has not been necessary to experiment on an actual piece of equipment, only to produce a specification with realistic values; the idealisations in the mathematical model of the system are only too easy to identify, we are more concerned with the appropriateness of the design methods when parameters do not have convenient values.

Section 2. The System.

The example considered is similar to one given by D'Azzo and Houpis (1966) and is shown in Fig. 5.1.

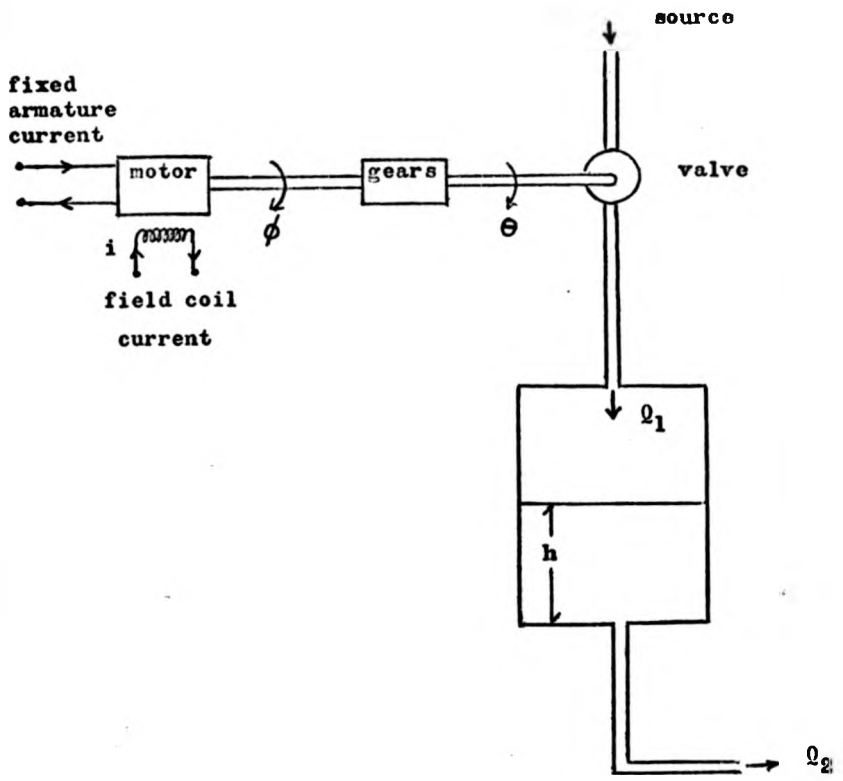


Fig. 5.1.

The system is designed to give a constant flow of water from the lower outlet of the tank and to remove, as far as possible, fluctuations in the flow from the source; it must also respond quickly to changes in the desired flow Q_2 . We shall assume that the system is governed by the following linearised equations.

Motor torque,

$$G = ci \quad ; \qquad (5.2.1.)$$

Moment of inertia of motor, shaft, gears and valve, measured at the motor = J; damping constant of this part of the assembly = B. Therefore

$$J\ddot{\phi} + B\dot{\phi} = G = ci \qquad (5.2.2.)$$

where ϕ is the angle turned through by the motor. Gear ratio = r so, if angle turned through by valve = θ

$$\theta = r\phi \; . \qquad (5.2.3.)$$

Inflow,

$$Q_1 = a\theta \qquad (5.2.4.)$$

Outflow,

$$Q_2 = bh \qquad (5.2.5.)$$

Cross sectional area of tank = A, hence

$$\dot{h} = \frac{1}{A}(Q_1 - Q_2) \; . \qquad (5.2.6.)$$

If equilibrium values are denoted by bars we have

$$\bar{Q}_1 = \bar{Q}_2 \quad , \qquad (5.2.7.)$$

so, from (5.2.4.) and (5.2.5.)

$$a\bar{\theta} = b\bar{h}$$

and from (5.2.3.)

$$\bar{\theta} = r\bar{\phi} \; ,$$

therefore

$$\bar{h} = ar\bar{\phi} / b \qquad (5.2.8.)$$

At equilibrium equations (5.2.1.) and (5.2.2.) imply that

$$\bar{\dot{\phi}} = 0 \ , \ \bar{i} = 0 \ , \ \bar{G} = 0 \ . \qquad (5.2.9.)$$

The differential equations (5.2.2.) and (5.2.6.) can now be framed in state space terms with the states defined as departures from the means of h, ϕ and $\dot{\phi}$. Set

$$x_1 = h - \bar{h}$$
$$x_2 = \phi - \bar{\phi}$$
$$x_3 = \dot{\phi} - \bar{\dot{\phi}} = \dot{\phi}$$

and the control

$$u = i - \bar{i} = i \ ,$$

so

$$\dot{x}_1 = \frac{1}{A} (- b x_1 + a r x_2)$$
$$\dot{x}_2 = x_3$$
$$\dot{x}_3 = -\frac{B}{J} x_3 + \frac{c}{J} u \ .$$

In matrix form these equations become

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -b/A & ar/A & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -B/J \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ c/J \end{bmatrix} u \qquad (5.2.10.)$$

that is

$$\dot{x} = A x + B u$$

This procedure is very straightforward, but if dealing with a real system the approximations must be kept well in mind. The major sources of error here are the linear relationships assumed to hold between outflow and depth and also between inflow and valve opening. The former is only true if the flow in the outlet pipe is laminar, which could be possible but we should then be limited to relatively low rates of flow. The latter is unlikely to be true with most common types of valve but it must be remembered that we are often concerned with relatively small fluctuations about the mean, so these non-linearities will then have comparatively small effect. The mathematical model of the system could be made more complex by allowing for the inductance of the field coil, but if it is assumed that the time constant of this coil, inductance divided by resistance, is small compared with the time constants of the rest of the system then this omission is justified. Damping is rarely linear and in this case will probably be a combination of Coulomb friction, viscous effects and eddy currents. However, the qualitative effect of damping oscillations is the same and in this sort of problem it is mainly the form of the response, rather than the strict numerical results, that we are interested in. Obviously a closer examination would reveal further defects in the mathematical representation, but the main factors to be aware of have already been discussed.

We should like to set up a realistic example without involving ourselves in technical problems of valve and motor specifications; this can be achieved by choosing some absolute

physical variables, to set the scale of the system, and determining all other parameters in terms of time constants.

It will be assumed that the tank has a cross sectional area of $2m^2$ and when the valve is half open an equilibrium flow rate of $0.1\ m^3 s^{-1}$. If the corresponding equilibrium depth of water is $1m$ and the time the tank would then take to empty at this flow rate equals $20s$ then

$$\bar{h} = 1\ m$$

$$a = 0.03183\ m^3 s^{-1} rad^{-1}$$

$$b = 0.1\ m^2 s^{-1}.$$

The motor will be assumed to give an output power of 30 W at 5000 r.p.m. when the control current is at its maximum value of 3A. If, in the absence of damping, it takes 2 s to reach this speed at full power and also at this speed takes 3s to turn the valve from fully closed to fully open (1 revolution) then

$$c = 0.019\ N\ m\ A^{-1}$$

$$J = 0.000328\ kg\ m^2$$

$$r = 0.006$$

$$\bar{\phi} = 523.6\ rad.$$

The damping constant B can be determined by assuming that if the control current is switched off the motor's angular velocity halves every 2 s, then

$$B = 1.133 \times 10^{-4}\ Nms$$

physical variables, to set the scale of the system, and determining all other parameters in terms of time constants.

It will be assumed that the tank has a cross sectional area of $2m^2$ and when the valve is half open an equilibrium flow rate of $0.1\ m^3 s^{-1}$. If the corresponding equilibrium depth of water is $1m$ and the time the tank would then take to empty at this flow rate equals $20s$ then

$$\bar{h} = 1\ m$$
$$a = 0.03183\ m^3 s^{-1} rad^{-1}$$
$$b = 0.1\ m^2 s^{-1} .$$

The motor will be assumed to give an output power of 30 W at 5000 r.p.m. when the control current is at its maximum value of $3A$. If, in the absence of damping, it takes $2\ s$ to reach this speed at full power and also at this speed takes $3s$ to turn the valve from fully closed to fully open (1 revolution) then

$$c = 0.019\ N\ m\ A^{-1}$$
$$J = 0.000328\ kg\ m^2$$
$$r = 0.006$$
$$\phi = 523.6\ rad.$$

The damping constant B can be determined by assuming that if the control current is switched off the motor's angular velocity halves every $2\ s$, then

$$B = 1.136 \times 10^{-4}\ Nms$$

physical variables, to set the scale of the system, and determining all other parameters in terms of time constants.

It will be assumed that the tank has a cross sectional area of $2m^2$ and when the valve is half open an equilibrium flow rate of $0.1\ m^3 s^{-1}$. If the corresponding equilibrium depth of water is $1m$ and the time the tank would then take to empty at this flow rate equals $20s$ then

$$\bar{h} = 1\ m$$
$$a = 0.03183\ m^3 s^{-1} rad^{-1}$$
$$b = 0.1\ m^2 s^{-1} .$$

The motor will be assumed to give an output power of 30 W at 5000 r.p.m. when the control current is at its maximum value of $3A$. If, in the absence of damping, it takes $2\ s$ to reach this speed at full power and also at this speed takes $3s$ to turn the valve from fully closed to fully open (1 revolution) then

$$c = 0.019\ N\ m\ A^{-1}$$
$$J = 0.000328\ kg\ m^2$$
$$r = 0.006$$
$$\bar{\phi} = 523.6\ rad.$$

The damping constant B can be determined by assuming that if the control current is switched off the motor's angular velocity halves every $2\ s$, then

$$B = 1.136 \times 10^{-4}\ Nms$$

Therefore the state equations are

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.05 & 9.549\times10^{-5} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -0.347 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 58.18 \end{bmatrix} u \qquad (5.2.11.)$$

Section 3. The Optimal Control.

If optimal control theory is to be applied to this problem it is necessary to specify a cost function. We shall limit ourselves to the standard quadratic form; it will appear later there is enough difficulty in choosing suitable parameters for this, so any further alternatives would make the amount of computation unmanageable. One other limitation that will be made is to take the period of integration to be infinite, a justifiable assumption as we are interested in driving the state variables right back to zero. Also, for practical reasons stated in Chapter 2, we should like the feedback gains to be time invariant and choosing this form for the cost function ensures that such a criterion is met.

The main concern with this system is to keep the outflow rate as close to its desired equilibrium value as possible. Since this flow rate is proportional to h one can infer that x_1 must be kept close to zero. This condition cannot be taken in isolation, it is essential to remember the constraints of the physical system. Though an electric motor may have a linear torque-current relationship over part of its range it will definitely experience saturation if the current is too heavy. The amplifier supplying the motor will also be subject

to saturation effects, so the cost function must penalise high values of control current. When using a quadratic cost function it is not possible to put a hard limit on the size of the control variables (one would have to use Pontryagin's maximum principle), but the greater the weighting given to this part of the cost the lower the absolute values of the control variables that result when the control system is operating. Hence for the example under consideration we choose

$$J = \int_0^\infty x_1^2 + R u^2 \, dt \qquad (5.3.1.)$$

where R is a positive scalar constant open to choice.

We now come to one of the greatest stumbling blocks in the application of control theory, a problem that I feel has never been satisfactorily resolved, that is choosing the parameters of the cost function. Although the cost function is very convenient mathematically and penalises what one feels intuitively should be penalised, it is very hard to quantify. In an example of this simplicity it is common to think in terms of the criteria of classical control theory, for example damping factor, gain and phase margins. It is difficult to say to what extent the formulation we are using here is lacking or whether it is simply a matter of inexperience in using optimal control theory for practical design. I think one must be prepared to use an iterative procedure: first use intuition and experience to specify the cost function, calculate the optimal control, look at some typical responses and then be prepared to modify the original cost function. Obviously increased experience will reduce the number of steps necessary.

It is all very well to say this is unsatisfactory and could be avoided by plotting a Nyquist chart, but what of the problem of high order system where there may be dozens of independently variable feedback gains?

Optimal control comes into its own here as the number of degrees of freedom can be greatly reduced. For example, look at the optimal culling policy example at the end of Chapter 4; there are technically an infinite number of feedback gains and even in the discretisation used there are 18 which must present a formidable task to the "classical" designer.

Returning to the example in hand one way of helping the intuitive process for choosing R is to use dimensionless variables. The elements of x and u can be expressed as fractions of their likely maximum values which can be of great help in deciding the relative weighting. This is equivalent to defining a new cost function

$$J' = \int_0^\infty \left[\frac{x_1}{x_{1\,max}}\right]^2 + \rho \left[\frac{u}{u_{max}}\right]^2 \, dt$$

where ρ is a dimensionless weighting factor. Since we are trying to minimise J the constant factor $1/(x_{1\,max})^2$ may taken outside the integral sign and then comparing the resulting expression with (5.3.1.) we get

$$R = \rho\left[\frac{x_{1\,max}}{u_{max}}\right]^2$$

In the example x_{1max} is the maximum likely deviation of the depth from the equilibrium which, if this level corresponds to the valve being half open, equals 1m. u_{max} has been specified as 3A, thus

$$R = \rho/9 \ .$$

We still have no really satisfactory way of choosing ρ or R a priori but in this relatively simple example it is possible to experiment with different values. The optimal control is quite simply calculated using the iterative method of Chapter 1 and it is then straightforward to look at the response of the system for any given initial state. In general, the lower the value of R the greater the control current fed to the motor, on the other hand, though, the water level is controlled in a shorter time. We wish to avoid the motor or amplifier being overloaded or the valve hitting its stops but the system must respond as quickly as possible. As mentioned earlier, optimal linear regulator theory cannot be completely successful in finding the "best" compromise, the designer still has to exercise his judgement. Despite its shortcomings the theory must not be rejected out of hand, even in this simple example it is an extremely useful aid. Its main advantage is that it gives a way of choosing the relative sizes of the three feedback gains, effectively reducing three degrees of freedom in the designer's choice to one, that is picking a value of R.

Table 5.1. shows the results of using the method of Chapter 1 to find the optimal control and as can be seen the lower R the higher the feedback gains. Although this method of finding the feedback gains can be made as accurate as desired it must be remembered that our knowledge of the system parameters is not particularly accurate when the linearising assumptions are taken into consideration. As a test of the sensitivity of the system we have calculated the optimal control for the case when the damping on the motor, one of the most difficult variables to ascertain, is zero; these results are also presented in Table 5.1. and the feedback gains plotted against R (log scale) are shown in Fig. 5.2. It can be seen that the damping makes very little difference to k_1 and k_2 but about halves k_3, which is not surprising as both the damping and k_3 relate torque on the shaft to the shaft's angular velocity. One other thing that is apparent from Table 5.1. is how good the iterative method of Chapter 1 is at finding the optimal control; all the feedback gains are accurate to one part in 10^4 and, although they vary considerably with R, the number of iterations necessary is nearly constant showing the procedure's insensitivity to the iteration starting point.

The responses of x_1, x_2 and u for the damped system are shown in Figs. 5.3. - 5.5. for various values of R; the initial state corresponds to the case in which we wish to change the equilibrium flow rate from $0.15\ m^3\ s^{-1}$ to $0.1\ m^3\ s^{-1}$ (i.e. reduce x_1 from 0.5m to 0m). Now, these

TABLE 5.1

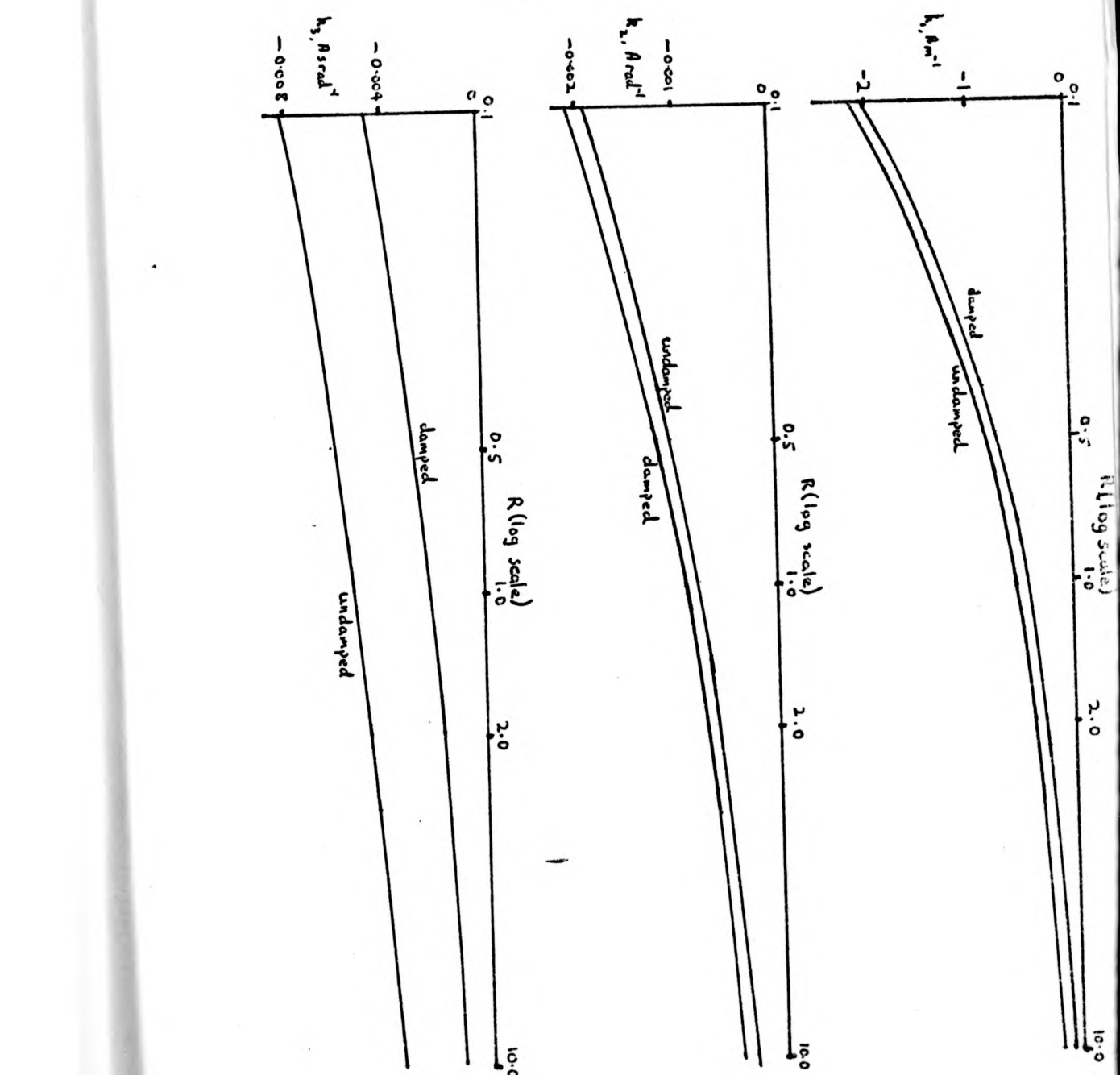
Optimal control results

R	UNDAMPED				DAMPED			
	k_1 Am^{-1}	k_2 $Arad^{-1}$	k_3 $Asrad^{-1}$	N	k_1 Am^{-1}	k_2 $Arad^{-1}$	k_3 $Asrad^{-1}$	N
0.1	-2.15500	-0.00192	-0.00813	12	-2.06067	-0.00210	-0.00442	12
0.5	-0.85741	-0.00106	-0.00605	13	-0.78506	-0.00120	-0.00280	12
1.0	-0.57066	-0.00082	-0.00531	13	-0.50804	-0.00094	-0.00227	13
2.0	-0.37713	-0.00063	-0.00465	14	-0.32399	-0.00073	-0.00183	13
10.0	-0.13964	-0.00034	-0.00341	14	-0.10653	-0.00040	-0.00106	14

All iterations start at (k_1 k_2 k_3) = (-0.5 -0.5 -0.5) and

k_1, k_2, k_3 converge to 0.01% in N iterations.

Fig. 5.2


k_1, Am^{-1}
0
-1
-2
0·1
0·5
1·0
2·0
10·0
R (log scale)
damped
undamped
k_2, A rad^{-1}
-0·001
-0·002
k_3, A rad^{-2}
-0·004
-0·008

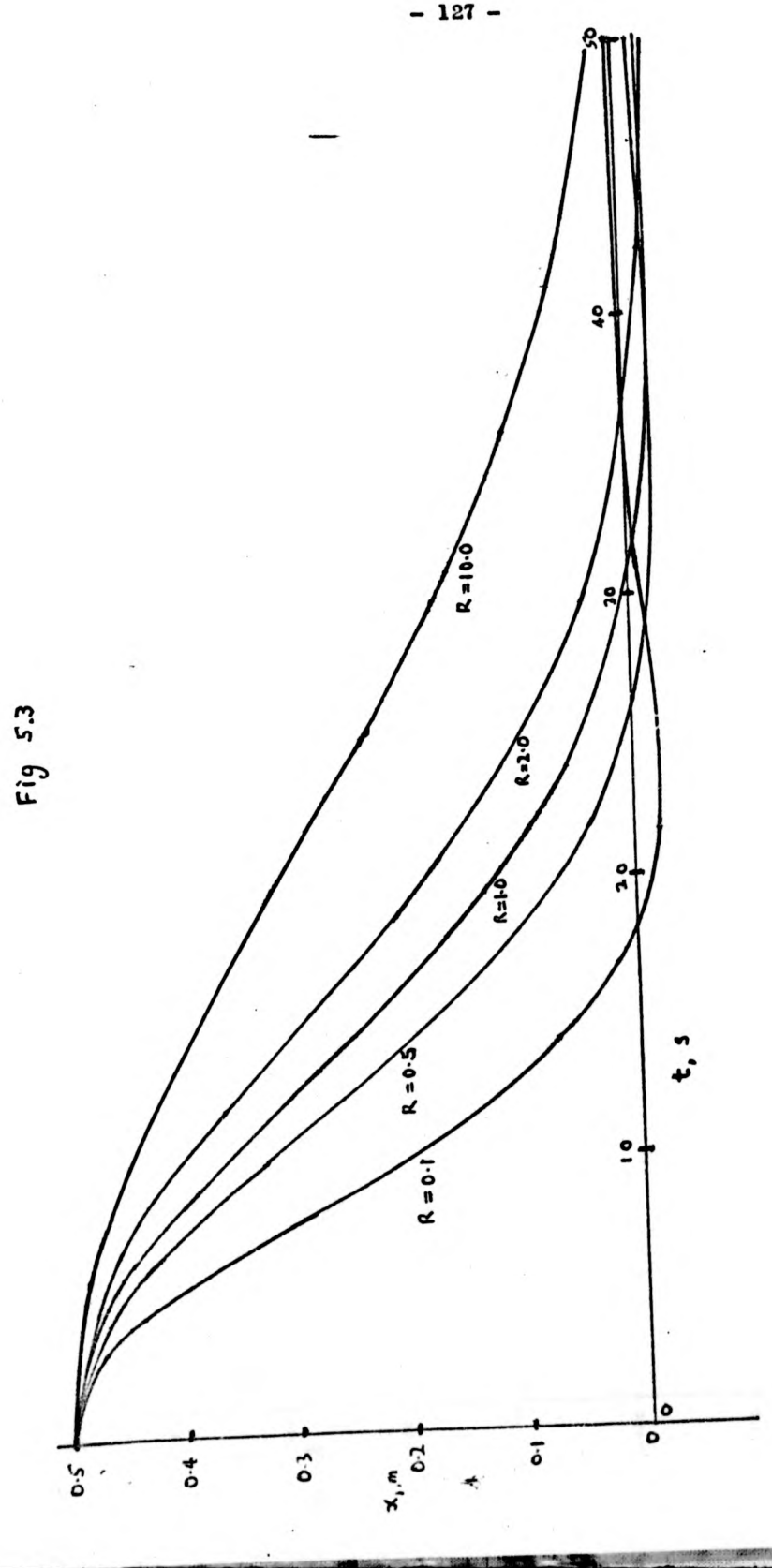


Fig 5.3

Fig 5.4

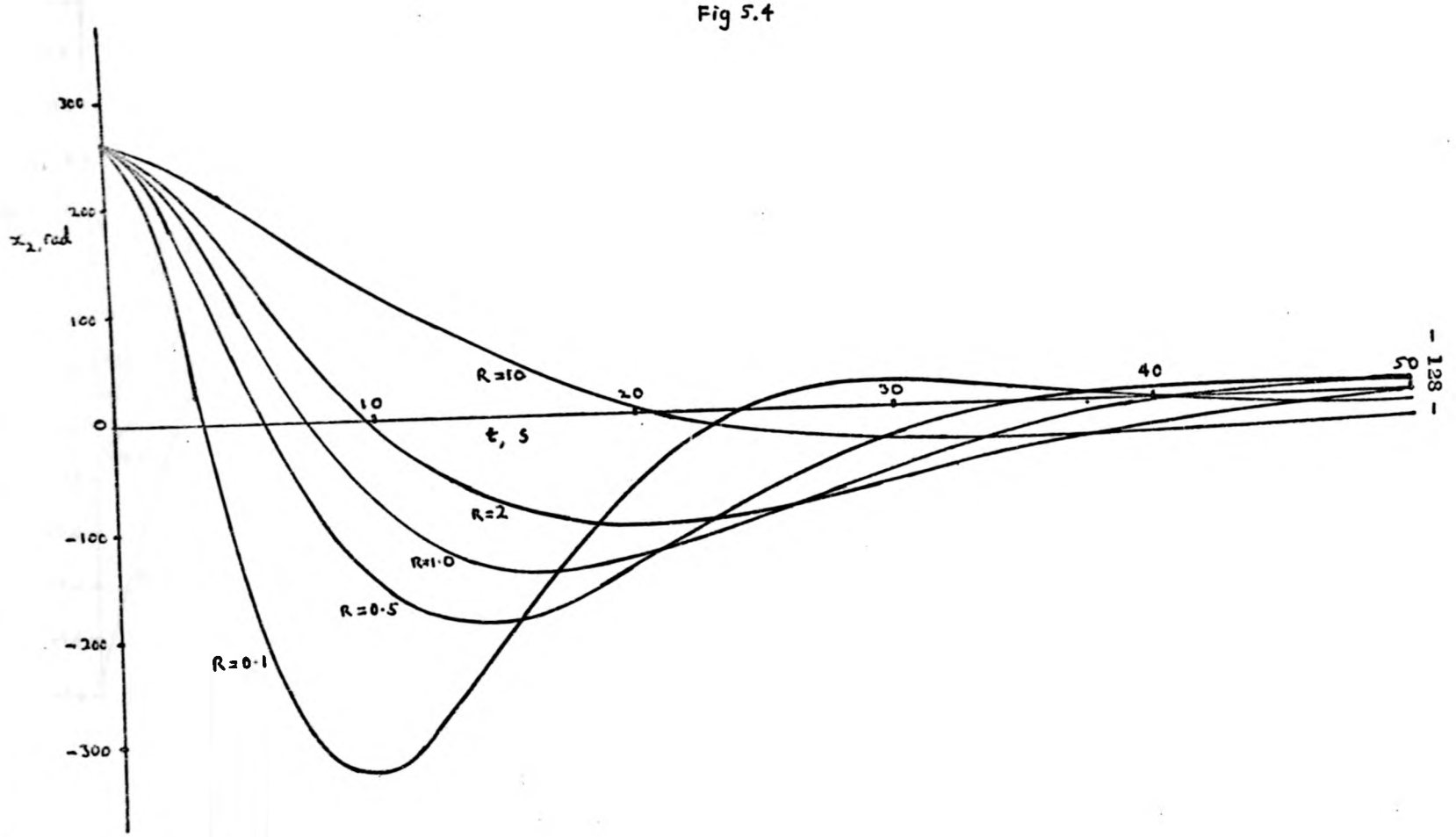
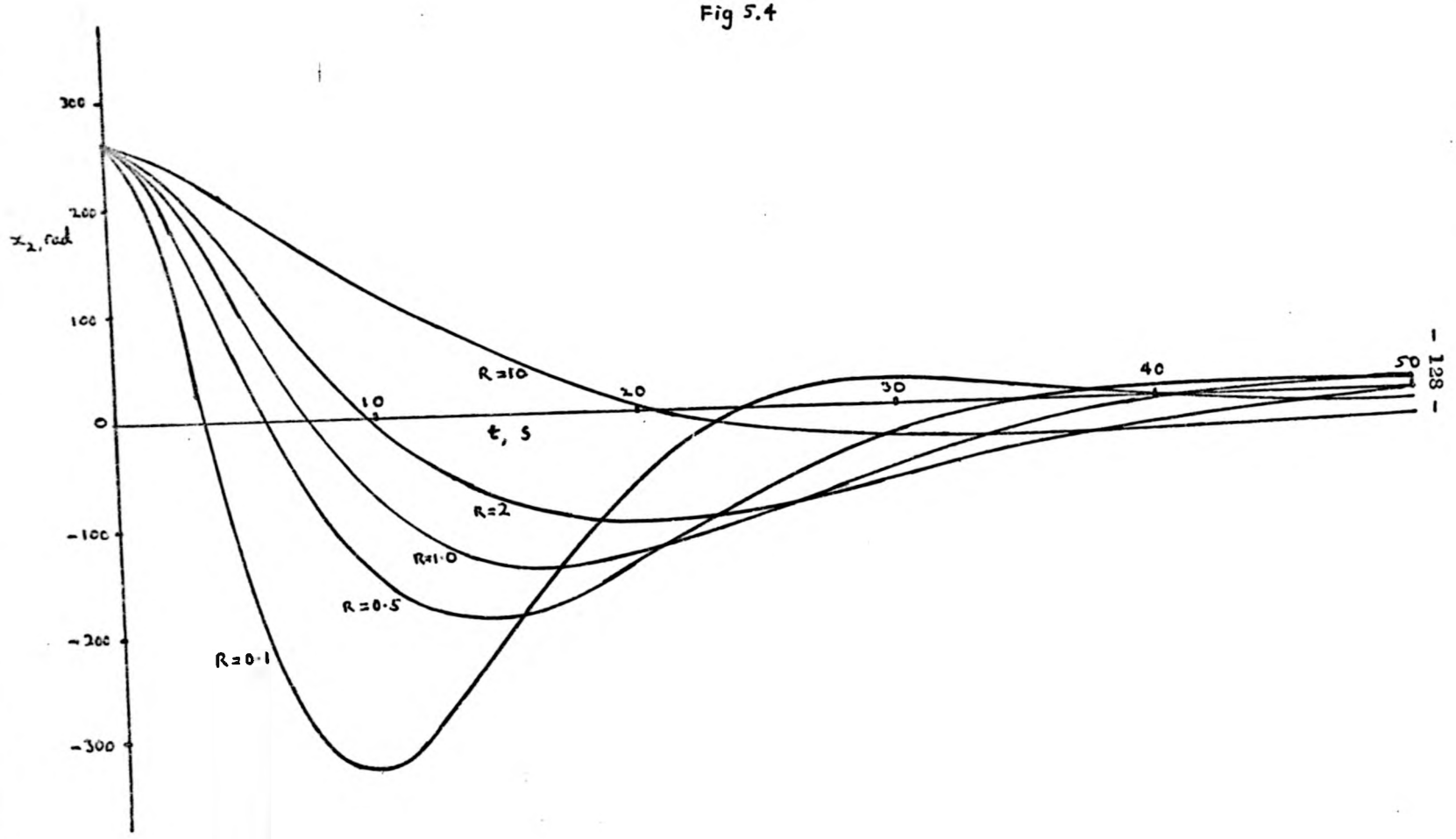

300
200
100
0
-100
-200
-300
x_2, rad
10
20
30
40
50
t, s
R=10
R=2
R=1·0
R=0·5
R=0·1

Fig 5.4


300
200
100
0
-100
-200
-300
x_2, rad
10
20
30
40
50
t, s
R=10
R=2
R=1·0
R=0·5
R=0·1

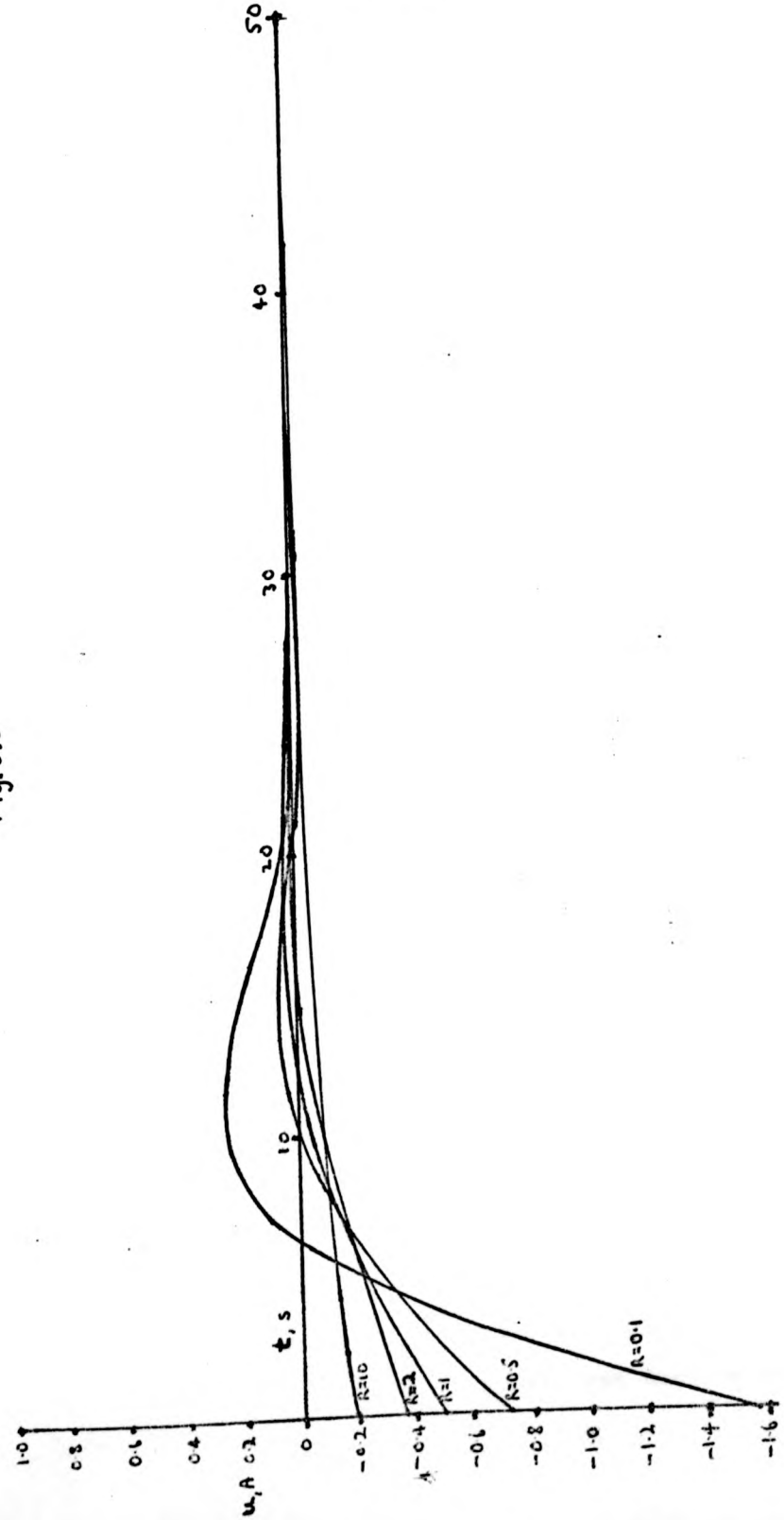

50
40
30
20
10
t, s
0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
1.0
0.8
0.6
0.4
0.2
u, A
R=10
R=2
R=1
R=0.5
R=0.1

Fig. 5.5

responses only represent one kind of problem that the system could be asked to deal with, producing a desired flow that can change with time, it will also have to reduce the effects of random fluctuations in the main supply; it is, in fact, these two aspects that make it desirable to use such a feedback control system in the first place. The uses to which the system is to be put determine which are the most important properties of the response, for example whether the increased overshoot that accompanies quicker reaction is acceptable. Even this simple example shows the many aspects of control design and how optimal control theory can be one extremely useful tool to aid the designer but is not the panacea to cure all his troubles.

Section 4. The Constrained Optimal Control.

Having considered the optimal control problem we shall examine the consequences of the angular velocity of the motor not being able to be measured. We shall assume that the level of the water and the position of the valve, rather than the motor, can be ascertained, that is the output vector $\mathbf{y}$ is given by

$$\mathbf{y} = C\mathbf{x}$$

where
$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.006 & 0 \end{bmatrix} \qquad (5.4.1.)$$

One great advantage of the optimal control is that it is necessarily stable, Ogata (1967), but once we restrict

the possible feedbacks there is no guarantee that the system can be stabilised. It is therefore very useful to find what combinations of the two feedback gains under our control give rise to a stable system. Since numerical search techniques will be used it is essential to know that our initial guess does lead to a finite value of the cost function. In this case we have

$$u = k_1 y_1 + k_2 y_2 = k_1 x_1 + 0.006\, k_2 x_2 \quad ,$$

from (5.4.1.), which can be used in (5.2.11.) to give

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -0.05 & 9.549\times10^{-5} & 0 \\ 0 & 0 & 1 \\ 58.18k_1 & 0.349k_2 & -0.347 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} . \qquad (5.4.2.)$$

The characteristic equation of the square matrix is found by setting

$$\det \begin{vmatrix} \lambda + 0.05 & -9.549\times10^{-5} & 0 \\ 0 & \lambda & -1 \\ -58.18k_1 & -0.349k_2 & \lambda + 0.347 \end{vmatrix}$$

to zero. On doing this one gets the following cubic equation

$$\lambda^3 + 0.397\lambda^2 + (0.0174 - 0.349k_2)\lambda - 0.00555k_1 - 0.0175k_2 = 0 \quad . \qquad (5.4.3.)$$

It is now possible to apply the Routh-Hurwitz criterion to this equation, D'Azzo and Houpis (1966), and obtain the following conditions for the system to be stable

$$\begin{aligned} &0.0174 - 0.349k_2 > 0 \\ &-0.00555k_1 - 0.0175k_2 > 0 \\ &0.397(0.0174 - 0.349k_2) > -0.00555k_1 - 0.0175k_2 \end{aligned} \qquad (5.4.4.)$$

It is obvious that if the 2nd and 3rd of these conditions are met then the 1st will follow automatically; the 2nd condition gives

$$k_2 < -0.317\, k_1$$

and the 3rd $k_2 < 0.0463\, k_1 + 0.0577$.

The unstable regions in the (k_1 , k_2) plane are shown shaded in Fig. 5.6. so the clear area gives the region in which both the above conditions hold. We are now in a position to use the methods of the earlier chapters on this constrained optimisation problem and hopefully a fair number of useful points will come up that will form part of a background of experience for future work in this field.

Section 5. Application of methods.

Before we plunge wholeheartedly into calculation of the optimal control according to various criteria it is worthwhile to see if the results of Chapter 2 can be used to set any bounds on the value of the cost function. We have here that Q is only positive semi-definite so it may be that the method is inapplicable. In the general formulae in Section 4 of Chapter 2 we have to substitute the particular forms for the matrices appropriate to this example, namely that R is scalar, B can be written

$$B = \begin{bmatrix} 0 \\ 0 \\ b \end{bmatrix} \qquad (5.5.1.)$$

Fig.5.6

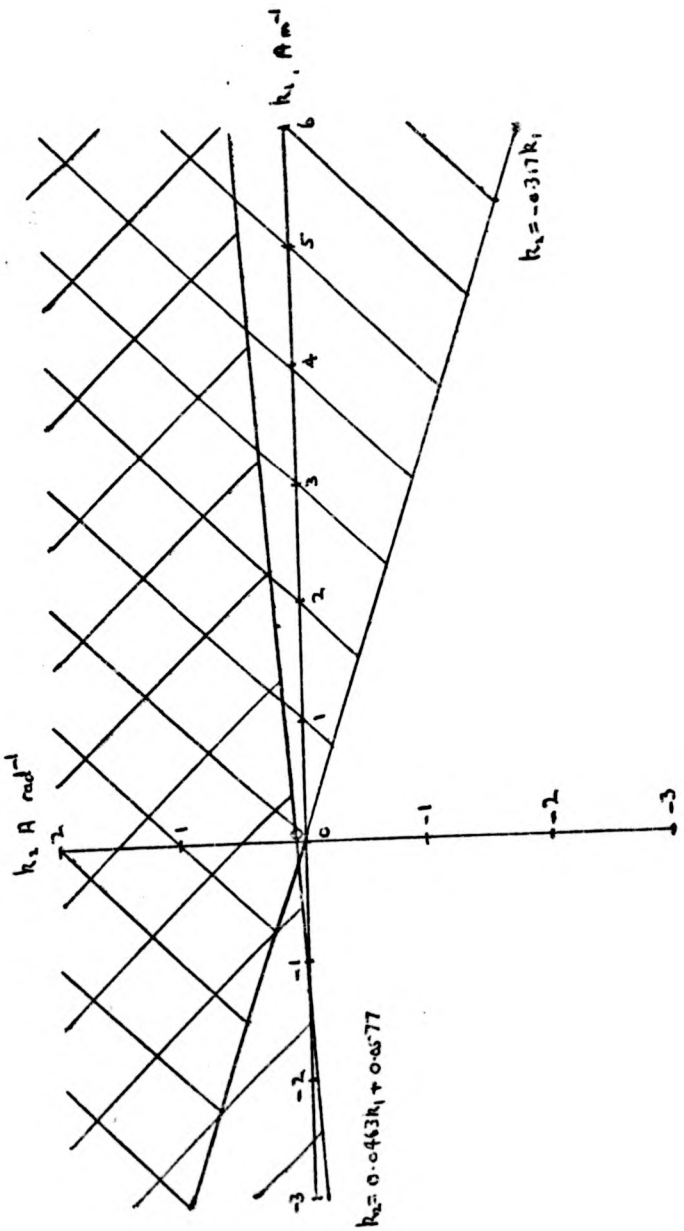
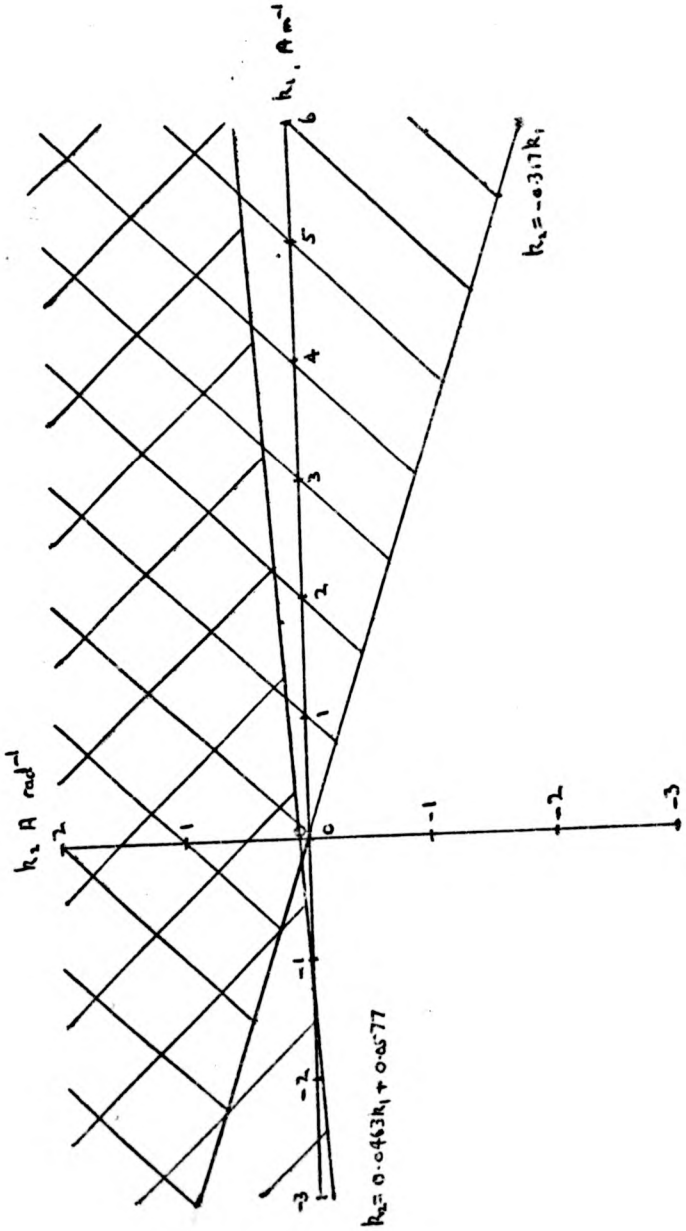

k_1, A rad^{-1}
k_2 A rad^{-1}
$k_2 = -0.3{\cdot}17k_1$
$k_2 = 0{\cdot}443k_1 + 0{\cdot}0{\cdot}77$
6
5
4
3
2
1
0
-1
-2
-3

Fig.5.6


k_1 A rad^{-1}
k_2 A rad^{-1}
$k_2 = -0{\cdot}317 k_1$
$k_2 = 0{\cdot}0443 k_1 + 0{\cdot}0577$
6
5
4
3
2
1
0
-1
-2
-3

and the control we wish to compare with any other

$$u_o = -\bar{R}^{-1}B^*P_o x = \begin{bmatrix} k_{1o} & \rho k_{1o} & 0 \end{bmatrix} x = K_o x \qquad (5.5.2.)$$

where $\rho = 0.006$. (2.4.18.) gives the condition that must be met by $\gamma > 0$ if we are to find a useful bound; writing this in finite dimensional terms with x replacing z gives

$$\gamma x'Qx + \frac{\gamma}{1+\gamma} x'PB\bar{R}^{-1}B'Px > x'(P-P_o)B\bar{R}^{-1}B'(P-P_o)x \qquad (5.5.3.)$$

Substituting (5.5.2.) into (5.5.3.) gives

$$\gamma x'Qx + \frac{\gamma}{(1+\gamma)R} x'PBB'Px > \frac{1}{R}\left[x'PBB'x - x'K_o'B'Px - x'PBK_o x + x'K_o'K_o x\right] . \qquad (5.5.4.)$$

P is a symmetric matrix that may be written

$$P = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_2 & p_4 & p_5 \\ p_3 & p_5 & p_6 \end{bmatrix} \qquad (5.5.5.)$$

so, using (5.5.1.)

$$PBB'P = b^2 \begin{bmatrix} p_3^2 & p_3p_5 & p_3p_6 \\ p_3p_5 & p_5^2 & p_5p_6 \\ p_3p_6 & p_5p_6 & p_6^2 \end{bmatrix} = b^2\, pp'$$

where $p = \begin{bmatrix} p_3 & p_5 & p_6 \end{bmatrix}'$.

Hence $x'PBB'Px = b^2\, x'pp'x$,

so $x'PBB'Px$ can be zero whenever x is orthogonal to p. $x'Qx$ is zero if $x = \begin{bmatrix} 0 & x_2 & x_3 \end{bmatrix}'$ therefore if

$$x = \begin{bmatrix} 0 & -p_6 & p_5 \end{bmatrix}' ,$$

or any scalar multiple of this, both $x'Qx$ and $x'PBB'Px$ will be zero. Consequently, as the right hand side of (5.5.4.) is positive semidefinite, it is impossible to find a suitable y, therefore this method cannot be used to give bounds on the value of the cost function.

We must therefore move on to using Jameson's equations to find the constrained optimal control. The first case to be considered is the simple one where the initial state is given and we shall take it to correspond to the problem of changing the equilibrium flow rate from one value to another, that is

$$x_0 = [1 \quad 563.6 \quad 0]'$$

There are three ways of using Jameson's equations and we shall apply them all and compare their relative merits.

Jameson himself thought it might be better to use the expression for the gradient of the cost function with respect to the feedback gains in an algorithm to find the minimum cost rather than to solve the equations directly. Therefore a steepest descent algorithm was programmed and applied to this problem. It is straightforward to show that if one makes a small change δk_{ij} in the feedback gain k_{ij} which is directly proportional to $-\frac{\partial J}{\partial k_{ij}}$ then the change in the space of feedback gains is along a line of steepest descent. Success is not guaranteed when using this procedure as it is in the method of Chapter 1, a great deal depends on the step length chosen. If it is too large one can well overshoot the

bottom of the "valley" and end up with a cost greater than before, while if the step is too small it can be wasteful of computing time. It is possible to try and put in an automatic step length routine but it is difficult to devise one suitable for all conditions. Bearing in mind the uncertainties in the data, convergence in the feedback gains of 1% was aimed at; this is still probably finer than is strictly warranted but one does want to make sure that a spurious convergence is not found. The results are shown in Table 5.2. and it can be seen that the results are not very satisfactory, convergence has proved to be difficult to obtain.

There are two most likely reasons why a steepest descent algorithm might not work. The shape of the multidimensional surface over which we are searching may mitigate against rapid convergence; it might be possible to jump from one part to another if there is more than one local minimum or else the minimum may be very flat making it easy to wander around the "valley" without finding the very bottom. Problems of this sort mainly arise when the step length is too large compared with the radius of curvature of the surface. The second possibility is basically a cause of the problems just mentioned. Steepest descent methods involve us thinking in Euclidean terms of distance between points in multidimensional space; all co-ordinates are visualised as "distances" whereas in fact they may be measured in different units from one another. For example, in the system under consideration, k_1 is measured in A m^{-1} and k_2 in A rad^{-1}. When the computer program is

TABLE 5.2

Steepest descent from Jamesen's equations

R	initial step	initial feedback gains		final feedback gains		N
0.1	0.01	-1.0	-1.0	-1.26	-0.272	101
0.1	0.01	-0.3	-0.1	-1.25	-0.271	101
0.5	0.01	-0.3	-0.1	-0.515	-0.167	11
0.5	0.01	-1.0	-1.0	-0.544	-0.168	40
1.0	1.0	-0.34	-0.18	—	—	—
1.0	1.0	0.0	-1.0	—	—	—
1.0	0.01	-1.0	-1.0	-0.367	-0.135	42
1.0	0.1	-1.0	-1.0	—	—	—
1.0	0.02	-1.0	-1.0	-0.367	-0.135	42
2.0	0.02	-1.0	-1.0	—	—	—
2.0	0.01	-1.0	-1.0	—	—	—
2.0	0.01	0.0	-1.0	—	—	—
2.0	0.01	-0.3	-0.1	-0.242	-0.108	17
2.0	0.01	-0.24	-0.11	-0.241	-0.108	10
2.0	0.01	-0.34	-0.18	-0.242	-0.108	21
10.0	0.01	-0.34	-0.18	-0.0043	-0.111	21
10.0	0.001	-0.34	-0.18	-0.101	-0.0554	11
10.0	0.001	-0.1	-0.05	-0.0845	-0.0621	101
10.0	0.01	-0.1	-0.05	-0.0905	-0.0595	101

N = no. of iterations, initial step is numerical Euclidean distance in (k_1, k_2) plane, — indicates no convergence.

Convergence criterion for all cases is 1%.

executed all concept of dimension disappears and everything is dealt with as a pure number. Hence it may well happen that the step sizes in each variable are inconsistent with the relative sizes of the variables in the physical system. It could be worthwhile to express the variables of the system in dimensionless terms by dividing by suitable scale factors. In the example under consideration 1m of output y_1 corresponds to π radians of output y_2 which numerically are of the same order of magnitude, therefore the program was implemented without the use of scaling factors. It would have been possible to experiment with scaling or to have tried to improve the algorithm in other ways to see if convergence could be improved, but since other methods turned out to be preferable, it was not thought to be worth the effort.

The next approach using Jameson's equations to be tried was that of direct iteration. This is the method commented on in Chapter 2 in which the cost matrix P and the matrix $W = \int_0^T x(t)x^T(t)dt$ are calculated from the feedback gain matrix K, then a new K is derived from P and W. There is no known proof that this method converges or that each step guarantees a reduction in the cost and in fact the numerical results about to be presented refute this hypothesis; they also show, though, that when it works this is a highly efficient method.

The results are presented in Table 5.3. and the intermediate steps of the iteration in the (k_1, k_2) plane are shown in Figs. 5.7. - 5.13. The main point to be noticed is that the process

TABLE 5.3

Direct iteration from Jameson's equations, given initial state

R	initial feedback gains		final feedback gains		
	Am^{-1}	$Arad^{-1}$	Am^{-1}	$Arad^{-1}$	N
0.1	-1	-1	—	—	—
0.1	-7	-7	—	—	—
0.2	-1	-1	—	—	—
0.2	-7	-7	—	—	—
0.5	-1	-1	—	—	—
0.5	-7	-7	—	—	—
1.0	-1	-1	-0.368	-0.135	20
1.0	-2	-2	-0.365	-0.135	16
2.0	-1	-1	-0.241	-0.108	13
2.0	-2	-2	-0.241	-0.108	17
5.0	-1	-1	-0.134	-0.080	9
10.0	-1	-1	-0.0838	-0.0625	6

All convergence to 1%. N= no. of iterations

— indicates failure to convorge.

does not converge for $R<1$, the iterations going into the unstable region. In reality when the feedback gains are such as to make the system unstable the cost becomes infinite; however, the program was written in such a way that the Liapunov matrix equation (2.3.6.) was solved whether the system was stable or not. Since this is just a linear equation in the elements of P it can in general be solved, though P will only be positive definite if the system is stable. Hence Jameson's equations can simply be thought of as an algebraic problem equivalent to solving

$$\xi = f(\xi) \qquad (5.5.1.)$$

where ξ is an N-vector of unknowns and f is some given vector function. The fact that only a certain region of has physical significance can be ignored during the computation, though any results must, of course, be checked to see that they do correspond with reality. The direct iteration method may be expressed in the form of (5.5.1.) as

$$\xi_{n+1} = f(\xi_n) \qquad (5.5.2.)$$

and it is a well known result of numerical analysis that for this process to converge

$$\left\| \frac{\partial f}{\partial \xi}(\xi_0) \right\| < 1 \qquad (5.5.3.)$$

where $\frac{\partial f}{\partial \xi}(\xi_0)$ is the matrix of partial derivatives evaluated at the root ξ_0 . It can be seen in (2.3.3.) that the expression for K has R^{-1} at the front so, for the case when R is scalar, we can see that $\left\| \frac{\partial f}{\partial \xi} \right\|$ is inversely proportional to R which explains why the method is unstable

for our example if R is small. This is rather an intuitive way of looking at the problem, the complete analysis is rather more complex as P is a function of R as well, but it is fair to say that if $\|R\|$ is small, $\|R^{-1}\|$ is large and the iteration scheme is less likely to be successful. The only possible exception occurs when Q = 0 as then the elements of P are linear functions of the elements of R so the effects on R^{-1} and P of small $\|R\|$ could well cancel out.

On examining Figs. 5.10. - 5.13. where this iteration method does work it appears that, as expected, the convergence is much quicker and less oscillatory for large R. In fact when R = 1 the iteration starting at $(k_1,k_2) = (-2,-2)$ actually goes into the unstable region, comes out again, and converges to the correct root. This is a fortuitous consequence of still calculating P from the Liapunov matrix equation even though the system is unstable. However, it must be said that for $R \geq 2$ the iteration is very successful as it converges on the correct root very rapidly for all the starting points tried.

These examples show that the direct iteration method is not entirely satisfactory and can easily break down. It certainly does not guarantee a reduction in cost as it can move in one step from the stable region, with finite cost, to the unstable region where the cost is infinite. Therefore we should like to find a method that is more reliable and

the fractional step algorith, described in Chapter 2, appears to meet this requirement. As far as is known this has never been proposed before so experience in its use is very limited, but when applied to the example under consideration it is undoubtedly the best method tried for solving the constrained optimal control problem. As mentioned before it is a simple modification to the direct iteration method derived from Jameson's equations, instead of going all the way to the new value of $\mathbf{K}$ indicated by (2.3.3.) one only goes a certain fraction towards it. The results are shown in Table 5.4 . and the corresponding trajectories in the (k_1,k_2) plane are shown in Figs. 5.7. - 5.13 . along with those from the direct iteration. The individual steps are shown if they are far enough apart to be clear, otherwise a continuous line is drawn. It can be seen that the oscillatory behaviour is eradicated and k_1 and k_2 home in on the optimal point in a very smooth way. Not unexpectedly the number of iterations can be greater than in the direct method and the most critical factor seems to be the starting point. It might be advisable to test a few values of $\mathbf{K}$ to see which gave the lowest cost, but if there are a large number of feedback gains such a direct search requires considerable computational effort. The guaranteed convergence to at least a local minimum will usually make it worthwhile to use the fractional step algorithm from any stabilising point, the computer time used in the extra steps is probably less than that necessary to carry out a provisional search.

TABLE 5.4

Fractional step algorithm, given initial state

R	α	f_{max}	initial feedback gains		final feedback gains		N
			Am^{-1}	$Arad^{-1}$	Am^{-1}	$Arad^{-1}$	
0.1	0.2	0.05	-1	-1	-1.267	-0.272	31
0.1	0.2	0.05	-1	-0.05	-1.268	-0.272	18
0.1	0.1	0.05	-3	-3	-1.238	-0.272	8
0.1	0.1	0.2	-3	-3	-1.268	-0.272	7
0.2	0.2	0.05	-1	-1	-0.890	-0.221	36
0.2	0.2	0.05	-1	-0.5	-0.890	-0.221	22
0.5	0.2	0.05	-1	-1	-0.544	-0.168	45
0.5	0.2	0.05	-1	-0.5	-0.543	-0.167	29
1.0	0.2	0.05	-1	-1	-0.366	-0.135	53
1.0	1.0	0.05	-1	-1	-0.365	-0.136	46
2.0	0.2	0.05	-1	-1	-0.242	-0.108	62
2.0	1.0	0.05	-1	-1	-0.242	-0.108	56
5.0	0.2	0.05	-1	-1	-0.135	-0.0795	76
5.0	1.0	0.05	-1	-1	-0.135	-0.0796	80
10.0	0.2	0.05	-1	-1	-0.0837	-0.0624	88
10.0	1.0	0.05	-1	-1	-0.0836	-0.0625	79

All convergence to 1%. N= no. of iterations,

α = fraction moved towards indicated gains,

f_{max} = maximum fractional change allowed per iteration.

Key to figures 5.7 to 5.13

The optimal point is shown by a circle ⊙

The boundary of the stability region is shown as a dotted line .....

The direct iteration method is depicted by a chain dotted line with diagonal crosses for the iteration points

The fractional step algorithm is represented by a solid line with vertical crosses for the iteration points if they are far enough apart to be clear

Fig.5.7

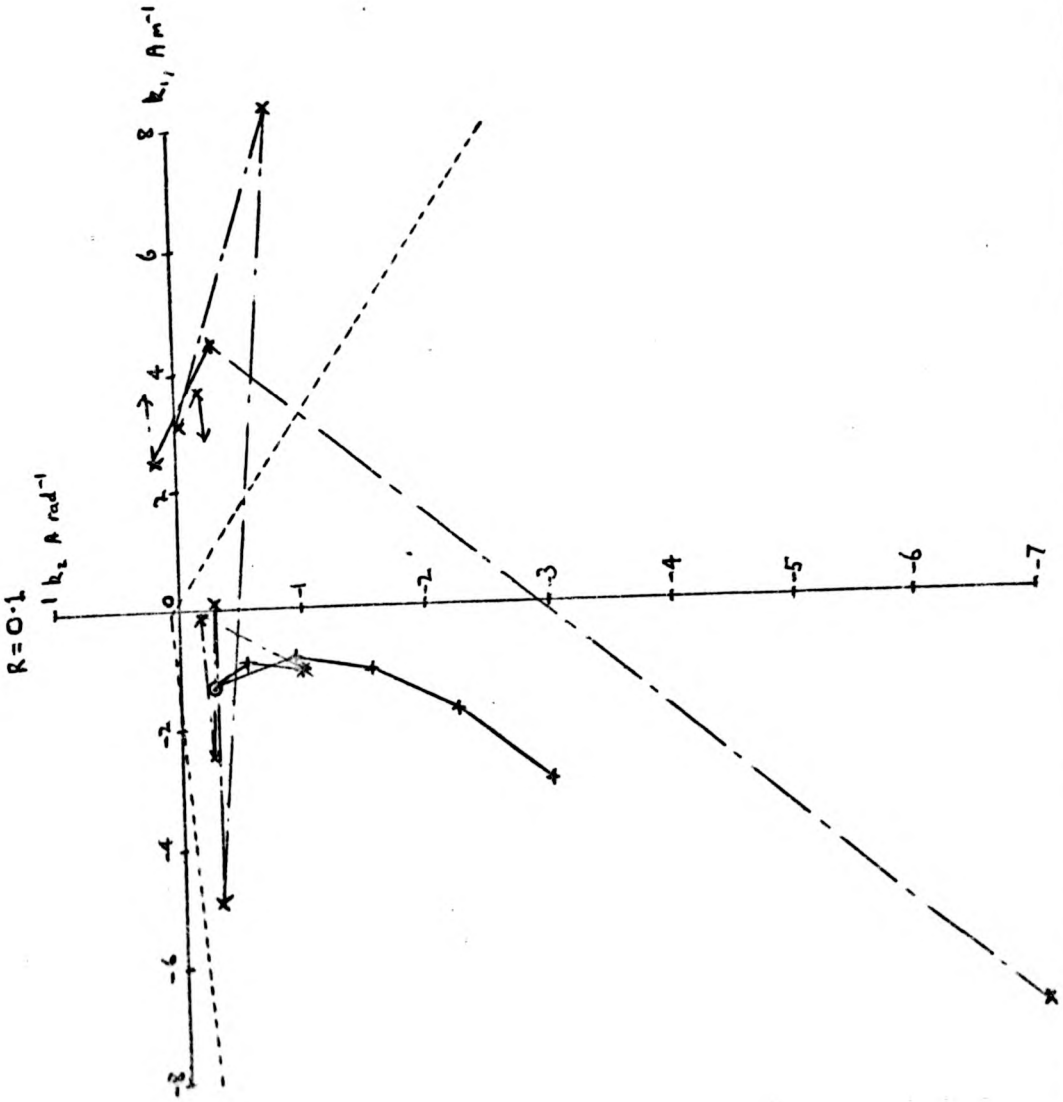


Fig.5.7

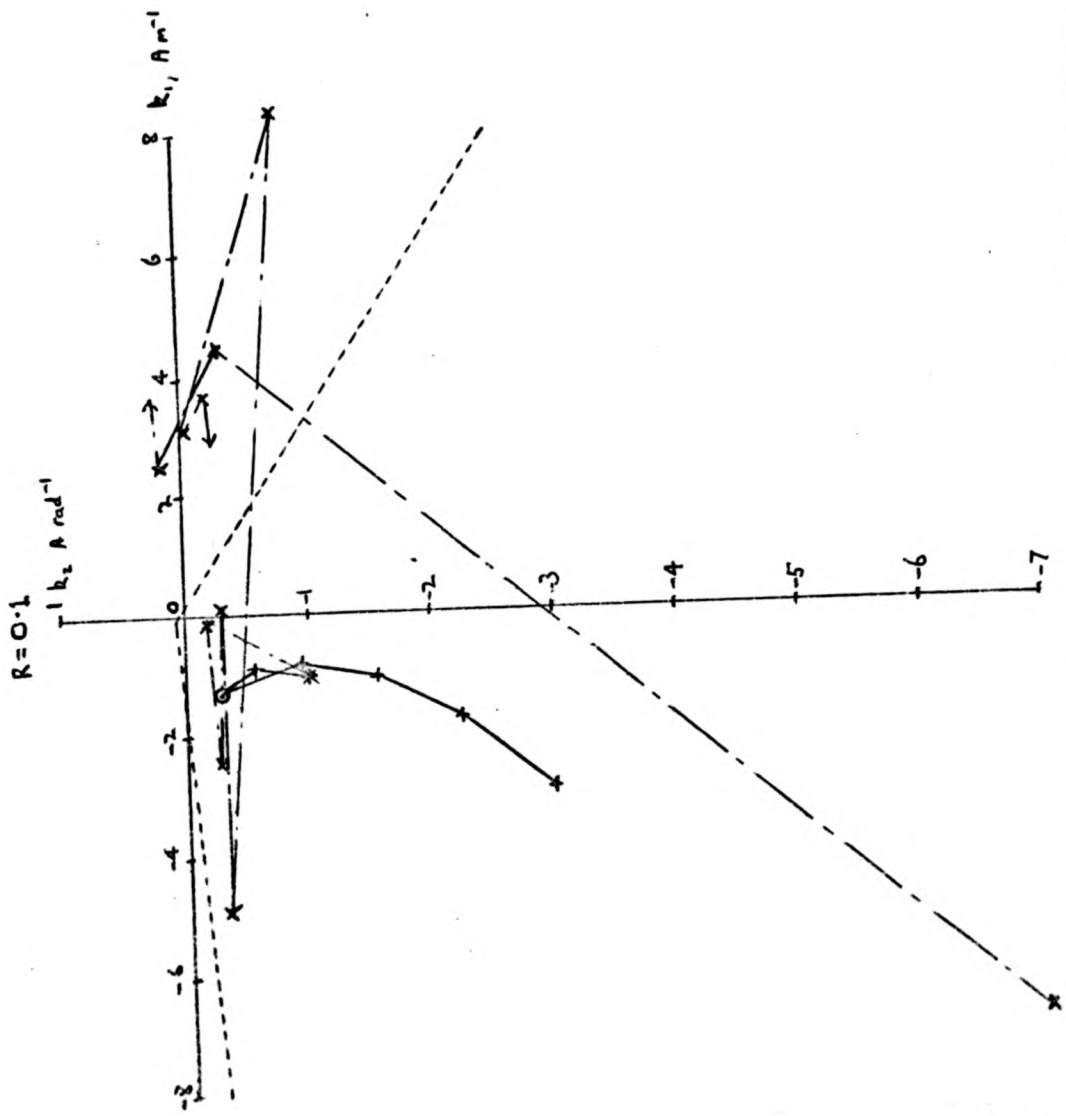


Fig.5.8

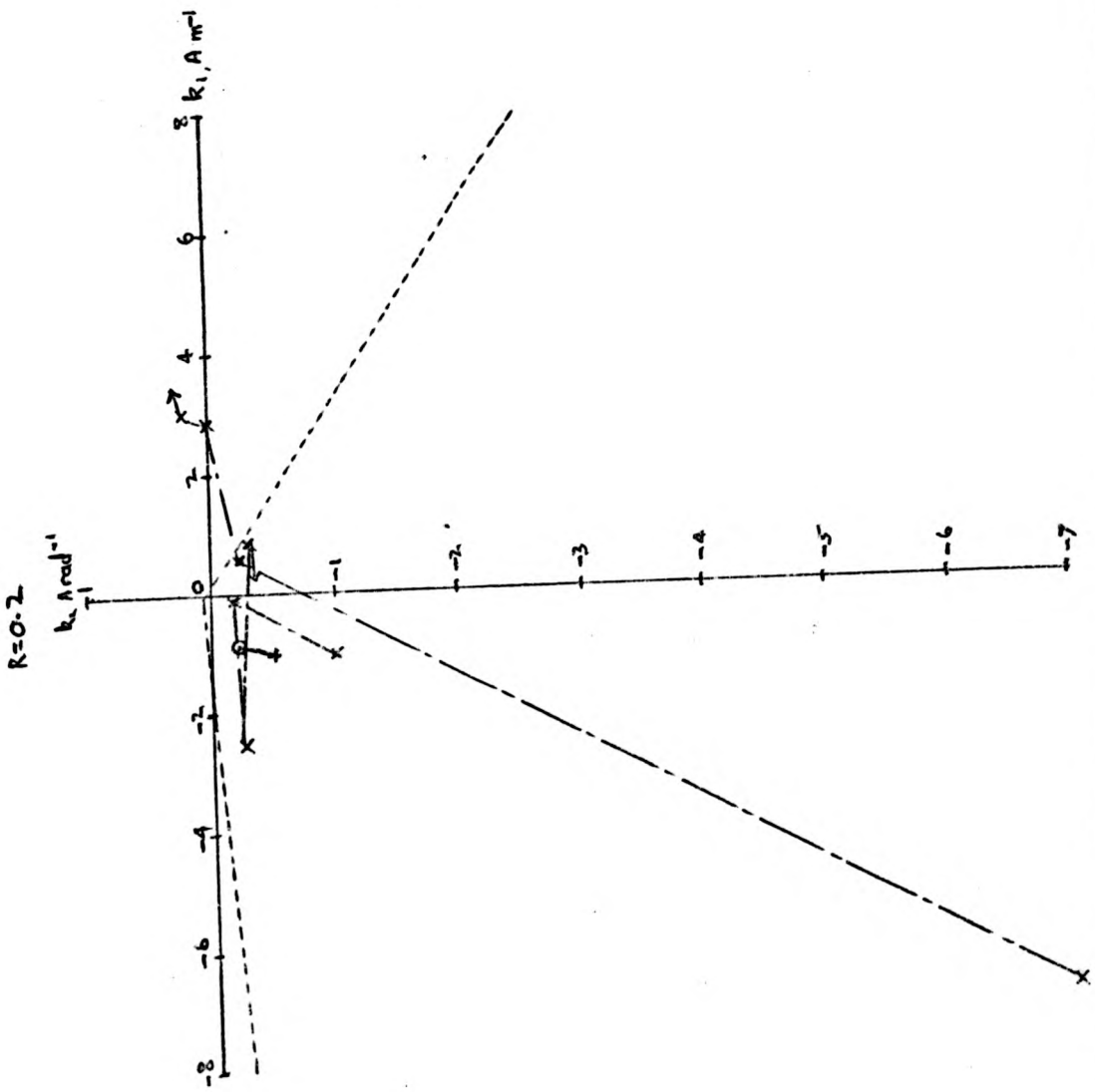


Fig.5.9

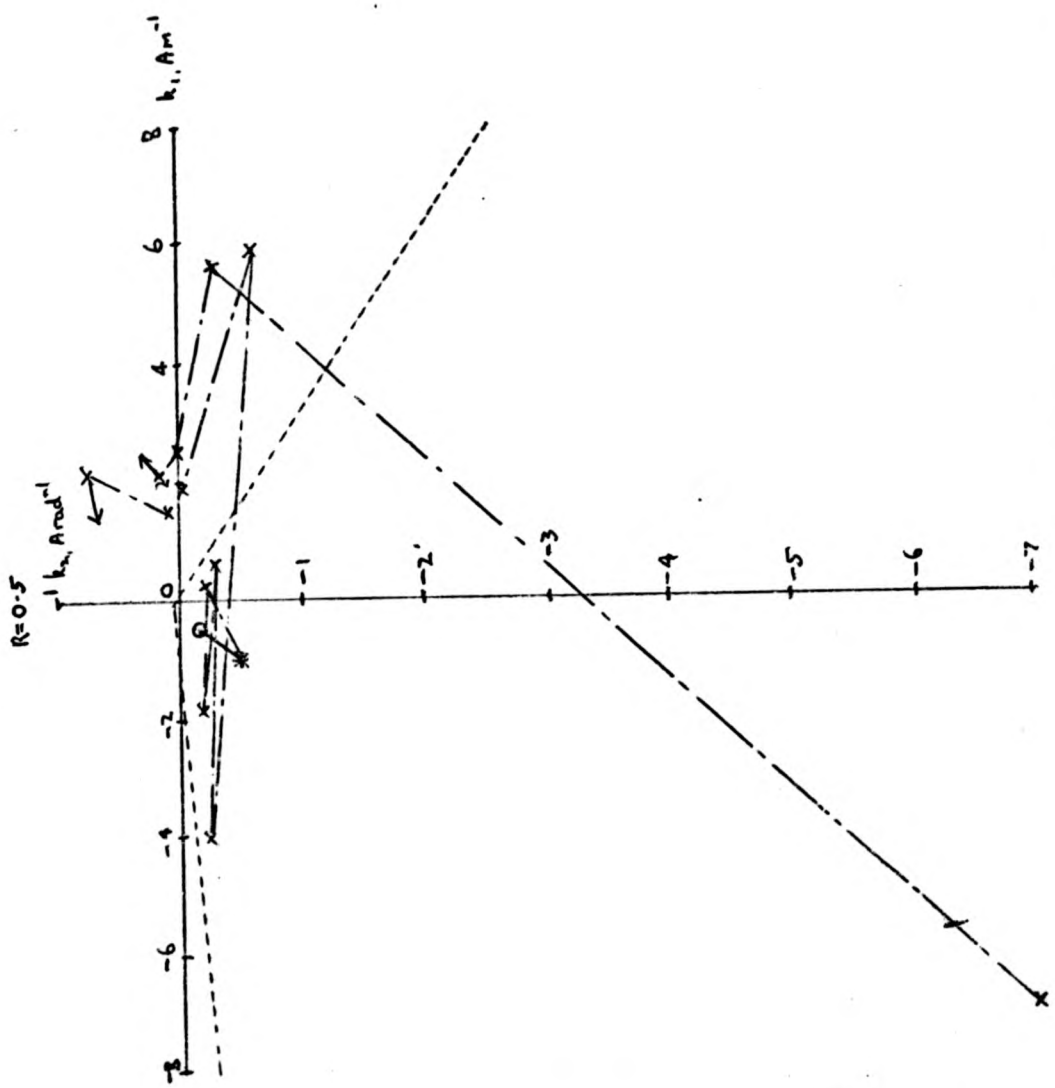


Fig. 5.10

R = 1·0

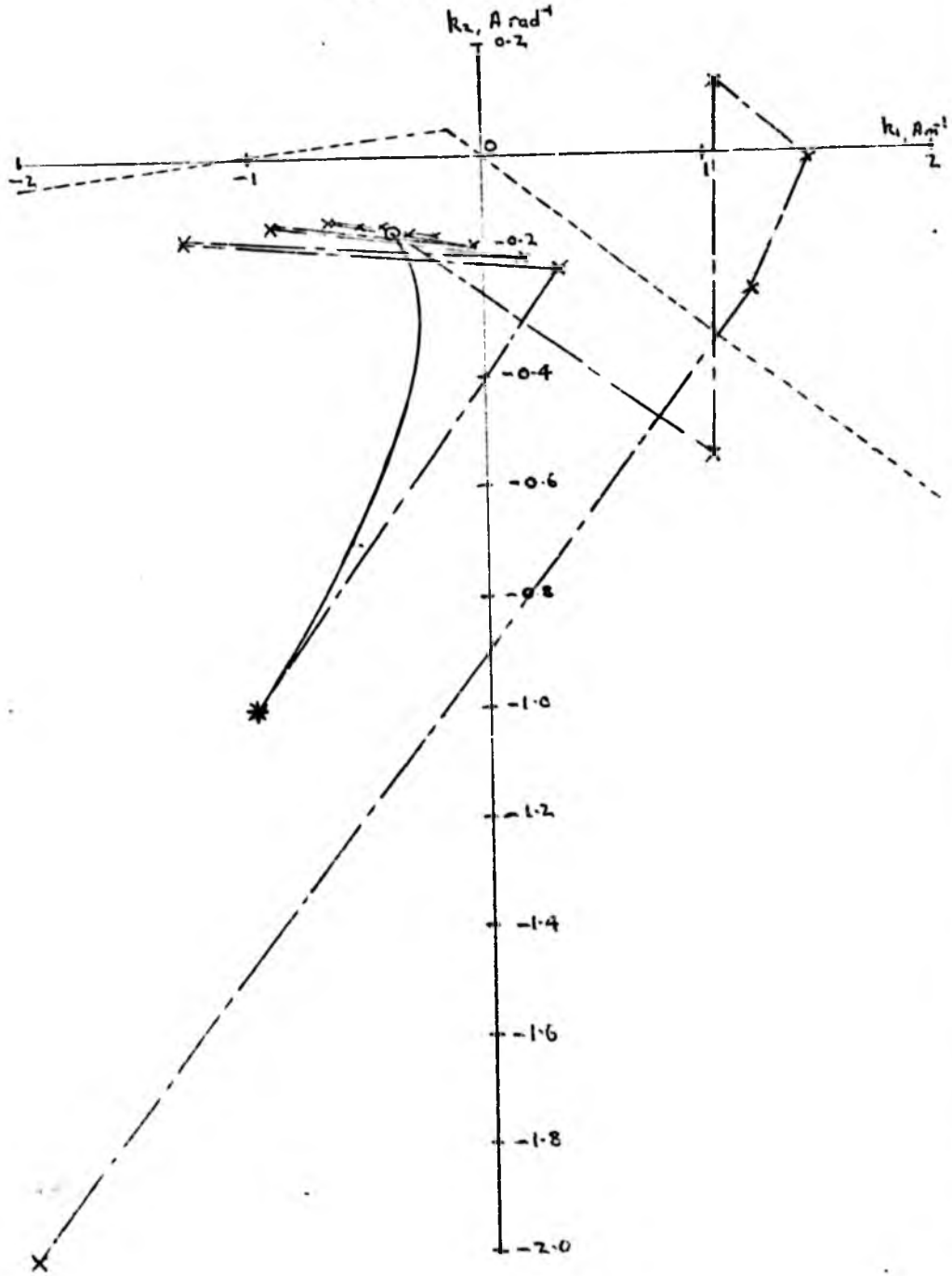


Fig.5.11

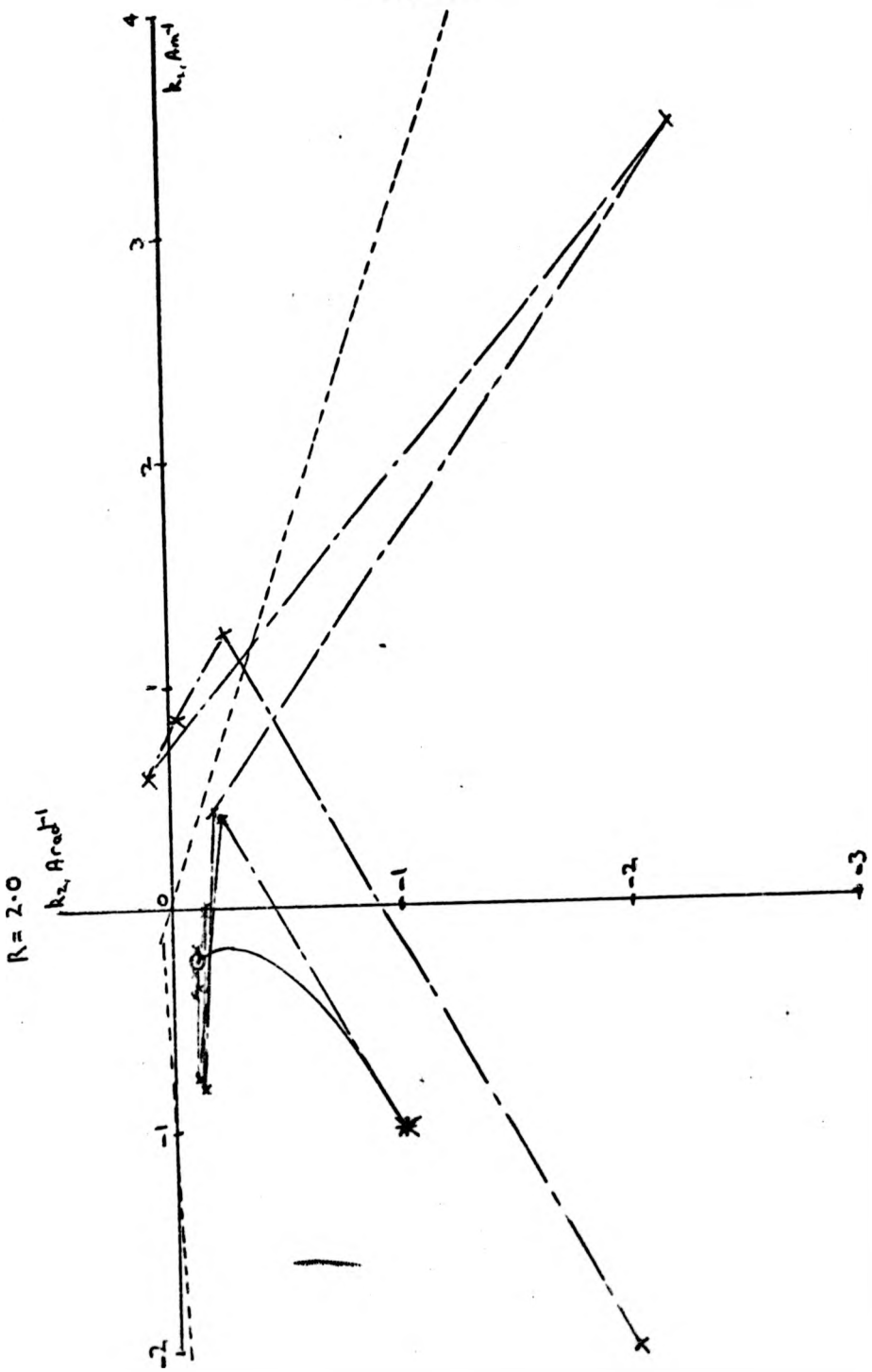


Fig.5.12

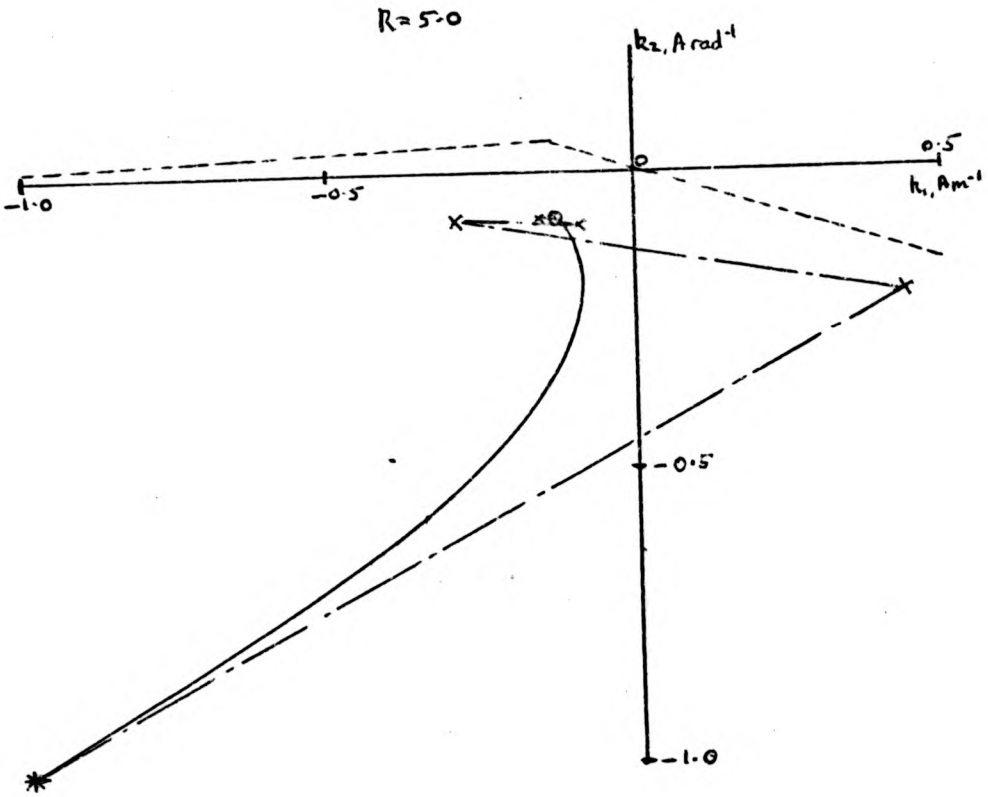
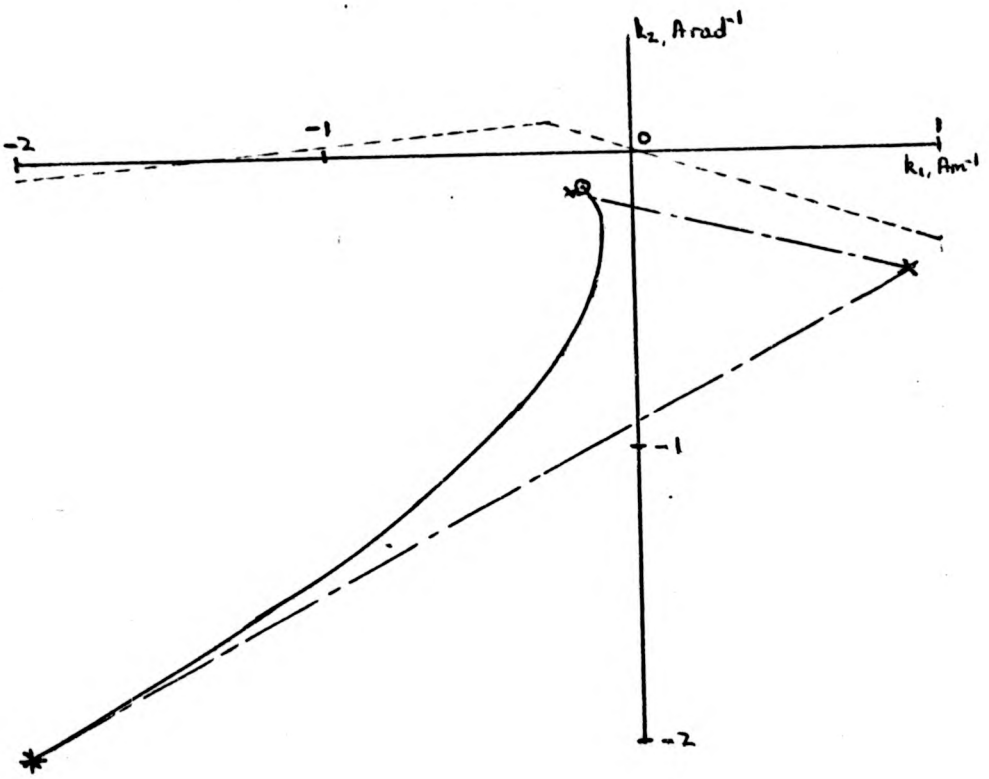


Fig.5.13

R=10·0



Unfortunately there does not seem to be a satisfactory a priori method of fixing the size of the fractional step, compromise has to be found between rapid convergence and reliability; the best procedure at the moment seems to be to set α high but reduce it on any particular step if there is going to be an element of K that varies by more than say 5%. As with any percentage based convergence criterion problems can arise if the result being approached is near zero. It is difficult to program a foolproof algorithm which will deal with all such cases and it is best just to keep this potential hazard in mind.

There is one more method that was tried for calculating the constrained optimal control problem and it is probably the most reliable and expensive of all. It consists of evaluating the cost over a grid in the (k_1,k_2) plane, making sure to remain in the stable region, and finding at which point on the grid the cost is lowest. If the minimum cost point is on the edge of the grid the search pattern is simply moved to centre on this point, if it lies within the grid the spacing between search points is reduced as well; the process is continued until the spacings are less than some convergence criterion. A 3 x 3 grid is not very satisfactory as it only has 1 interior point and 8 edge ones which means that the spacing is reduced quite infrequently. A 5 x 5 grid has 9 interior points and 16 on the edge which is an improvement, but evaluation of 25 values of the cost involves a considerable amount of computation, whereas the Liapunov matrix equation only has to be solved twice, for P and W, in the fractional step

method. However, if the search area starts off sufficiently large, convergence to the global optimal control is assured. Even searching over a two dimensional space as in this example is time consuming and if, say, there were 6 elements of $\mathbf{K}$ and the search was carried out over a hypercube of side 5 it would involve evaluating the cost $5^6 = 15625$ times which would be prohibitive. The attractiveness of this method depends very much on the accounting of one's computer expenditure. If the computer is self-owned, and the marginal cost low, then the direct search method is feasible; if one is paying a bureau on a machine usage basis then it is very likely to prove extremely expensive. Some runs were made using this method as a cross check, the same results shown in Table 5.5. were obtained but the computer time necessary was greater by about a factor of 10 compared with the iterative methods.

One other advantage of Jameson's equations is that they can also be used for the maximum eigenvalve criterion (3.2.5.), (3.2.6.). This involves finding

$$\min_{\mathbf{K}} \max_{\mathbf{x}_o} \frac{\mathbf{x}_o' \mathbf{P} \mathbf{x}_o}{\mathbf{x}_o' \mathbf{S} \mathbf{x}_o} \qquad (5.5.4.)$$

where $\mathbf{S}$ is some positive definite matrix. This is equivalent to finding

$$\min_{\mathbf{K}} \lambda_{\max}(\mathbf{P}\mathbf{S}^{-1})$$

and is only necessary to replace $\mathbf{x}_o$ by the eigenvector associated with the maximum eigenvalve of $\mathbf{P}\mathbf{S}^{-1}$. Finding

TABLE 5.5

Direct search

R	M	h_o	initial feedback gains		final feedback gains		N
			Am^{-1}	$Arad^{-1}$	Am^{-1}	$Arad^{-1}$	
1.0	3	0.2	-1	-1	-0.366	-0.136	18
1.0	5	0.1	-1	-1	-0.366	-0.136	13
1.0	5	0.01	-0.35	-0.14	-0.366	-0.135	4
2.0	5	0.01	-0.24	-0.11	-0.242	-0.108	6

All convergence to 1%. N = no. of search patterns, square grids, necessary for convergence. M = no. of points along each side of grid. h_o = size of grid element.

TABLE 5.6

Fractional step algorithm, minimax criterion

R	α	f_{max}	initial feedback gains		final feedback gains		N
			Am^{-1}	$Arad^{-1}$	Am^{-1}	$Arad^{-1}$	
0.1	0.2	0.05	-1	-1	-1.339	-0.259	32
0.2	0.2	0.05	-1	-1	-0.944	-0.210	36
0.5	0.2	0.05	-1	-1	-0.580	-0.159	42
1.0	0.5	0.5	-1	-1	-0.394	-0.129	8
2.0	0.5	0.5	-1	-1	-0.262	-0.103	9
5.0	0.5	0.5	-1	-1	-0.148	-0.0760	10
10.0	0.5	0.5	-1	-1	-0.0932	-0.0598	11

All convergence to 1%. See foot of Table 5.4 for definitions of N, α and f_{max}.

$$\max_{x_o} \quad \frac{x_o' P x_o}{x_o' S x_o}$$

is the same as finding the maximum of $x_o' P x_o$ subject to the constraint $x_o' S x_o = 1$, and since the elements of x_o have different dimensions we have to choose suitable scaling factors as the elements of S. The simplest way of doing this is to make

$$x_o' S x_o = \sum \left(\frac{x_{oi}}{x_{si}} \right)^2 \qquad \text{or}$$

$$S = \text{diag}\left(\frac{1}{x_{s1}^2}, \frac{1}{x_{s2}^2} \ldots\ldots \right)$$

where x_s is a vector of scaling factors. In the example under consideration this was chosen as

$$x_{s1} = 1 \text{ m}$$

$$x_{s2} = 523.6 \text{ rad}$$

$$x_{s3} = 523.6 \text{ rad s}^{-1} .$$

The fractional step algorithm was then used and the results are shown in Table 5.6. It can be seen that choosing a large step, with a 5% upper limit, has resulted in very quick convergence. The actual results are not that different from those obtained using a fixed initial state and considering the difficulty in precisely defining the cost function there is not much to be said between the two approaches, and it comes down again to the use one is going to put the system to. Must it perform as well as possible in a standard situation or be able to deal with unexpected conditions?

Section 6. Conclusions.

We have investigated a system which has been made as realistic as possible without actually taking measurements on physical apparatus and within the confines of using a linear model. The objective has been to design a controller for this system using the theory developed in the thesis with a view to seeing what problems arise. This is especially important since many examples used in theoretical work are extremely simple and chosen to illustrate points of the theory rather than to show how practical systems can be improved. It turns out in our example that the greatest problem confronting the designer is choosing the parameters of the quadratic cost function. The penalties thus defined cannot reflect all his targets with respect to speed of response, overshoot, damping factor etc but have only an imprecise aim which endeavours to somewhat reduce all the undesirable aspects of the response. However, there can be so many possible combinations of feedback gains that the unifying of the objectives into one cost function provides an extremely useful way of reducing the degrees of freedom of the design problem.

Once the quadratic cost function is specified the calculation of the totally observed optimal control is relatively straightforward, the method of Chapter 1 being extremely efficient. In fact, the greatest computational difficulty arises in solving the Liapunov matrix equation at every step; it is a linear equation in the elements of the upper triangle of a symmetric matrix and careful programming is required to transform into vector form so

Section 6. Conclusions.

We have investigated a system which has been made as realistic as possible without actually taking measurements on physical apparatus and within the confines of using a linear model. The objective has been to design a controller for this system using the theory developed in the thesis with a view to seeing what problems arise. This is especially important since many examples used in theoretical work are extremely simple and chosen to illustrate points of the theory rather than to show how practical systems can be improved. It turns out in our example that the greatest problem confronting the designer is choosing the parameters of the quadratic cost function. The penalties thus defined cannot reflect all his targets with respect to speed of response, overshoot, damping factor etc but have only an imprecise aim which endeavours to somewhat reduce all the undesirable aspects of the response. However, there can be so many possible combinations of feedback gains that the unifying of the objectives into one cost function provides an extremely useful way of reducing the degrees of freedom of the design problem.

Once the quadratic cost function is specified the calculation of the totally observed optimal control is relatively straightforward, the method of Chapter 1 being extremely efficient. In fact, the greatest computational difficulty arises in solving the Liapunov matrix equation at every step; it is a linear equation in the elements of the upper triangle of a symmetric matrix and careful programming is required to transform into vector form so

that standard simultaneous equation solving subroutines can be used.

When the information about the state is limited the most interesting points that come to light are those concerning the methods necessary for finding the constrained optimal control. Direct use of Jameson's equations proves unsatisfactory; simple iteration between the equations can go completely wrong and certainly does not guarantee a reduction in cost as was hoped. Also using the expression for the derivatives in a steepest descent algorithm is not very successful, without sophisticated automatic step length routines and scaling of the variables convergence is difficult to achieve. The best approach seems to be the fractional step method based on Jameson's equations; although this can take a fair number of steps to converge it is very robust, invariably finding a local minimum whatever the starting point. It nearly always works out in practice that a slow method which is certain to give correct results will be cheaper in computer costs than one which, though potentially efficient, can fail completely. If one uses a minimax criterion for this example, assuming that the initial state is not known, the optimal feedback controller is not radically altered. Since the cost function does not have a precise physical significance it is difficult to rank meaningfully controllers that lead to it having similar values.

The overall impressions gained from the exercise in this chapter are that there is considerable difficulty in defining the cost function to fit in with one's intuitive ideas of what is a good response. However, for all but the simplest systems, there is a very great advantage in using optimal control theory primarily for determining the relative sizes of the feedback gains. Similar conclusions can be drawn for partially observed systems but there is also the added difficulty of finding suitable numerical methods, even the best of these require appreciably more computing time than for the totally observed optimal control.

---oOo---

REFERENCES

1. M. Athans (1970) — "Towards a practical theory for distributed parameter systems". IEEE Transactions on Automatic Control, Vol. 15, No. 2, pp 245-274.

2. A.J. Pritchard & M.J.E. Mayhew (1971) — "Feedback from discrete points for distributed parameter systems". Int. J. Control, 1971, Vol. 14, No. 4, pp 619-630.

3. K.T. Parker (1970) — "Stability and optimal control of a distributed parameter system with feedback from discrete points". M.Sc. dissertation in Engineering Control, University of Warwick, 1970.

4. K. Ogata (1967) — "State space analysis of control systems". Prentice-Hall, 1967.

5. W.L. Brogan (1968) — "Optimal control theory applied to systems described by partial differential equations". Advances in Control Systems, Vol. 6, 1968. (Academic Press).

6. R. Bellman (1957) — "Dynamic Programming". Princeton University Press, 1957.

7. R.E. Kalman (1960) — "Contributions to the theory of optimal control". Bol. Soc. Mat. Mexicana, 5, pp 102-119.

8. L.S. Pontryagin, V.G. Boltyanski, R.V. Gamkredidze, E.F. Mishchenko (1962) — "The mathematical theory of optimal control processes". J. Wiley, New York, 1962.

9. M. Athans & P.L. Falb (1966) — "Optimal control: an introduction to the theory and its applications". McGraw Hill, New York, 1966.

10. P.K.C. Wang (1964) "Control of distributed parameter systems". Advances in control systems, 1964, pp 75-172. (Academic Press).

11. M. Kim & H. Erzberger (1967) "On the design of optimum distributed parameter systems with boundary control". IEEE Transactions on Automatic control, Vol. 12, No. 1, February 1967, pp 22-28.

12. J.L. Lions (1971) "Optimal control of systems governed by partial differential equations". Springer, 1971.

13. A.J. Pritchard (1972) "The linear, quadratic problem for systems described by evolution equations". University of Warwick, Control Theory Centre, Report No. 10.

14. A. Jameson (1970) "Optimization of linear systems of constrained configuration". Int. J. Control, 1970, Vol. II, No. 3, pp 409-421.

15. A.V. Balakrishnan (1965) "Optimal control problems in Banach spaces". J. Siam Control, Ser. A, Vol. 3, No. 1, 1965.

16. R. Curtain & A.J. Pritchard (1975) "The infinite dimensional Riccati equation for systems defined by evolution operators". J. Siam control, to be published.

17. D.G. Luenberger (1966) "Observers for multivariable systems". IEEE Transactions on Automatic Control, Vol. 11, April 1966, pp 190-197.

18. J.J. Bongiorno & D.C. Youla (1968) "On observers in multivariable control systems". Int. J. Control, 1968, 1970, Vol. 8, No. 3, pp 221-243.

19. J.J. Bongiorno & D.C. Youla (1970) "Discussion of 'On observers in multivariable control systems'". Int. J. Control, 1970, Vol. 12, No. 1 pp 183-190.

20. Z.V. Rekasius (1967) — "Optimal Linear regulators with incomplete state feedback". IEEE Transactions on Automatic Control, Vol. 12, June 1967, pp 296-299.

21. W.S. Levine, T.L. Johnson & M. Athans (1971) — "Optimal limited state variable feedback controllers for linear systems". IEEE Transactions on Automatic Control, Vol. 16, No. 6. December 1971, pp 789-793.

22. J.L. Douce (1963) — "An introduction to the mechanics of servomechanisms". English University Press (1963).

23. J.J. DAzzo & C.H. Houpis (1966) — "Feedback control system analysis and synthesis". McGraw-Hill, 1966.

24. J.G. Zieglor & N.B. Nichols (1942) — "Optimum settings for automatic controllers". Trans. ASME, 64:759, 1942.

25. M. Athans (1971) — "On the design of P-I-D controllers using optimal linear regulator theory". Automatica, 1971, 7, pp 643-647.

26. K.T. Parker (1972) — "Design of proportional-integral-derivative controllers by the use of optimal-linear-regulator theory". Proc. IEE, Vol. 119, No. 7, July 1972.

27. A.J. Pritchard & K.T. Parker (1974) — "A lower bound for the cost functional for control problems in Hilbert space". J. Inst. Maths Applics (1974) 13, pp 97-106.

28. A.J. Pritchard & K.T. Parker (1971) — "Simplified Lyapunov matrix equation with applications to the control of distributed parameter systems". Electronics Letters, 15th July 1971, Vol.7 No.14.

29. K.Yosida (1965) "Functional Analysis". Springer, 1965

30. R.Courant & D.Hilbert (1963) "Methods of mathematical physics". Interscience, 1963.

31. T.Kato (1966) "Perturbation theory for linear operators". Springer, 1966.

32. P.H.Leslie (1945) "On the use of matrices in certain population mathematics". Biometrika 35, pp 183-212.

33. J.R.Beddington & D.B.Taylor "Optimum age specific harvesting of a population". Biometrics, Vol 29, No.4, December 1973, pp801-809.